ARI Working Papers

Fort Huachuca Field Unit

1987-1991

July 2001

United States Army Research Institute for the Behavioral and Social Sciences

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy) July 2001	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) 1987-1991
4. TITLE AND SUBTITLE ARI Working Papers: Fort Huachuca Field Unit, 1987-1991		5a. CONTRACT OR GRANT NUMBER
		5b. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Akman, A., Knapp, B.G., Burnstein, D.D., Fichtl, T.C., Thompson, J.R., Lawton, G.W., Leddo, J.M., Cohen, M.S., Marshall, P.H., Zaklad, A.L., Moan, K.L., and Zachary, W.W.		5c. PROJECT NUMBER
		5d. TASK NUMBER
		5e. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences ATTN: TAPC-ARI-PO 5001 Eisenhower Avenue Alexandria, VA 22333-5600		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Avenue Alexandria, VA 22333-5600		10. MONITOR ACRONYM ARI
		11. MONITOR REPORT NUMBER WP Huachuca

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

ARI working papers were originally unofficial documents intended for limited distribution to obtain comments. These working papers are being archived in order to preserve material that was not included in other ARI publications. The material contained herein may not meet ARI's usual scientific or professional standards for publication.

14. ABSTRACT *(Maximum 200 words)*:

Nine working papers dealing with job analysis, military intelligence, Morse Code, image analysis, expert knowledge systems, nonverbal communication, and tactical deception.

15. SUBJECT TERMS

Job analysis, military intelligence, Morse Code, image analysis, expert knowledge systems, nonverbal communication, tactical deception, JCAT, MOS, Terra Scout

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON (Name and Telephone Number)
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited	172	David W. Witter (703) 617-0324

Fort Huachuca Field Unit Working Papers

Akman, A., & Knapp, B.G. (1990). The Job Comparison and Analysis Tool (JCAT): A method for matching MI MOS capabilities and IEW system demands. WP HUA 90-02.

Burnstein, D.D. (1991). Intelligence production as a model for processing information. WP HUA 91-02.

Fichtl, T.C., & Thompson, J.R. (1988). Measurement and evaluation of Military Intelligence (MI) Performance: Supplemental report. WP HUA 88-03.

Knapp, B.G. (1989). The role of the human operator in the Morse Code collection cycle: A descriptive model. WP HUA 89-03.

Lawton, G.W. (1988). Human factors and human performance issues in softcopy image analysis. WP HUA 88-01.

Lawton, G.W. (1989). Suggested Terra Scout experiment plan. WP HUA 89-02.

Leddo, J.M., & Cohen, M.S. (1987). A cognitive science approach to elicitation of expert knowledge. WP HUA 87-13.

Marshall, P.H. (1990). An evaluation of tests of sensitivity to nonverbal communication. WP HUA 90-03.

Zaklad, A.L., Moan, K.L., Zachary, W.W., & Knapp, B.G. (1988). A framework for tactical deception. WP HUA 88-04.

Working Paper

WP HUA 90-02

The Job Comparison and Analysis Tool (JCAT): A Method for Matching MI MOS Capabilities and IEW System Demands

Allan Akman, Akman Associates, Inc., and Beverly G. Knapp, ARI Field Unit-Ft. Huachuca

July 1990

Reviewed by: JULIE A. HOPSON
Chief, ARI Field Unit-Ft. Huachuca

Approved by: JULIE A. HOPSON
Chief, ARI Field Unit-Ft. Huachuca

Cleared by: ROBIN L. KEESEE
Director, Systems Research Laboratory

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

THE JOB COMPARISON AND ANALYSIS TOOL (JCAT): A METHOD FOR MATCHING MI MOS CAPABILITIES AND IEW SYSTEM DEMANDS

CONTENTS

List of Figures

List of Tables

THE JOB COMPARISON AND ANALYSIS TOOL (JCAT): A METHOD FOR MATCHING MI MOS CAPABILITIES AND IEW SYSTEM DEMANDS

Introduction

The U.S. Army Intelligence Center and School (USAICS) frequently must make decisions that involve matching Military Intelligence (MI) Military Occupational Specialties (MOS) to Intelligence/Electronic Warfare (IEW) systems. To develop methods for addressing critical MI MOS-IEW issues, the circumstances in which decisions are made about matching people and equipment must be understood. Army planners determine IEW system demands in a very different setting than that in which MOS capabilities are assessed. Whereas the latter is based on a personnel pool with relative stability which has been or can be measured in terms of job performance attributes, no similar pool of equipment exists when the focus is on new equipment. In fact, at the stage when IEW-MOS match issues should first be addressed, often very little definitive information exists about the equipment. The major opportunities to address the impacts of new IEW systems exist during its development and acquisition. The Army's Life Cycle Systems Management Model (LCSMM) and its Manpower and Personnel Integration (MANPRINT) program provide a programmatic framework in which IEW system demands must be addressed.

IEW System Acquisition

The LCSMM is the Army's formal process by which new IEW equipment is designed, developed, and eventually fielded. Army Regulation (AR) 70-1 establishes the Army's basic policies and regulations with regards to systems acquisition policy and procedures. USAICS has combat development proponency formulating doctrine, concepts, materiel requirements, and objectives which lead to IEW systems acquisition.

The system acquisition process involves five phases, each of which is separated from the succeeding phase by a major decision, or milestone, at which the appropriate DOD or Army authority reviews the developmental work accomplished and determines whether there is a basis for committing resources to proceed with the next phase. The phases, shown in Figure 1, include: mission area analysis, concept exploration and definition, concept demonstration and validation, full-scale development and low rate initial production (LRIP), and production and deployment/ operational support.

Within this formal setting, new IEW equipment is acquired. To match MI MOS capabilities with IEW system demands, several

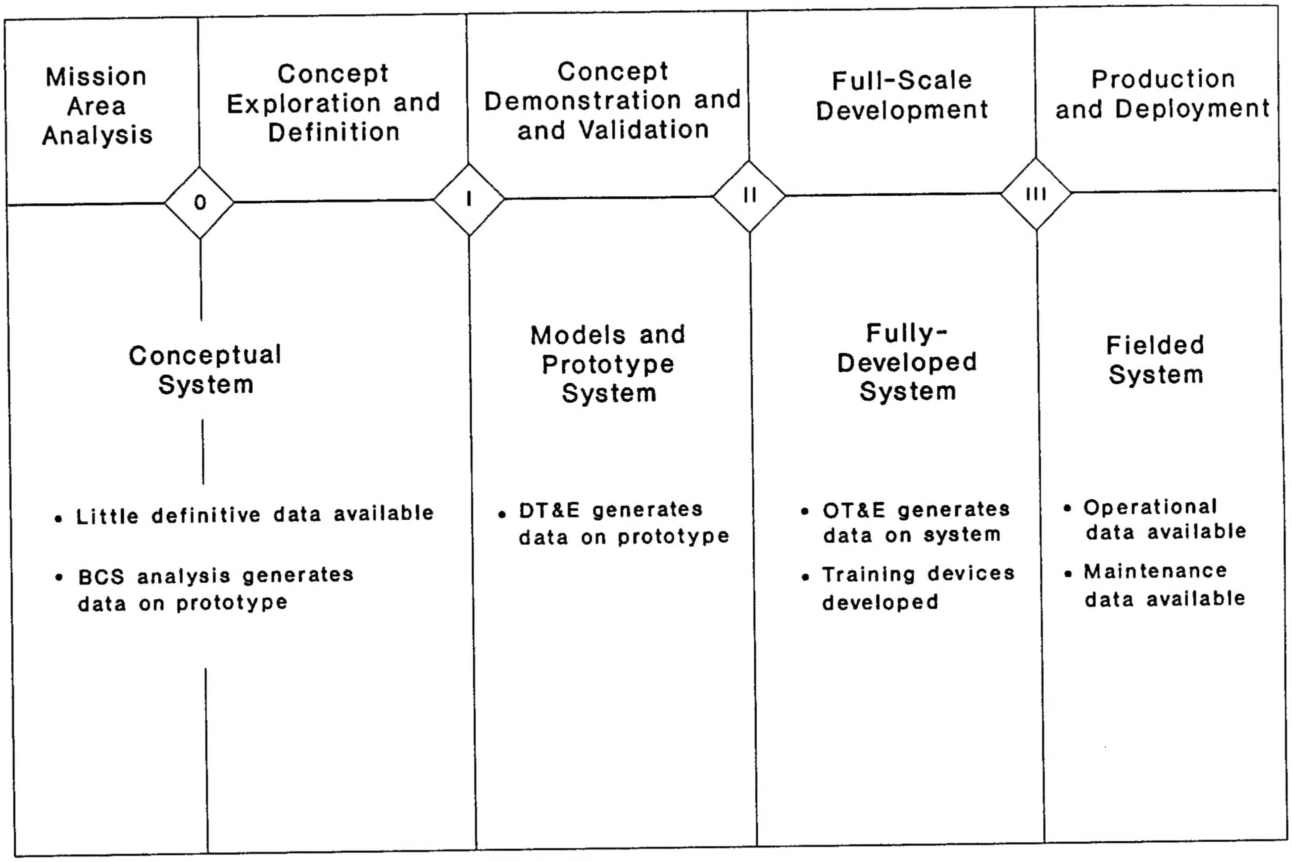


Figure 1. Life Cycle System Management Model

features of the LCSMM must be recognized. First the equipment exists only as concepts; there may be predecessor systems but in the front-end of the LCSMM there is no hardware or software. Prior to Milestone I, when this analysis should first occur, relatively little definitive data exist about the equipment. MOS analysis is based on a personnel pool in a relatively stable job environment. Therefore, MOS capabilities can be measured in terms of job performance attributes. No equivalent reference exists for new IEW systems analysis. The analyst generally must create a baseline comparison system (BCS) to initiate the analysis.

Milestone I, the decision to proceed with concept demonstration and validation, is considered by many observers as a demarcation point in characterizing the availability of information, the uncertainties, and the commitment to life cycle costs. Analyses before Milestone I are accomplished often with very little, if any, data, uncertainties are great, and, once decisions are made, most of the life cycle resource requirements are locked in. Although the analytical and decision framework existing prior to Milestone I is quite difficult, that is the timeframe in which to deal with manpower, personnel, and training (MPT) issues. Therefore, unlike analysis of MOS capabilities which is on-going in nature and not time constrained, analysis of IEW demands using BCSs must begin early in the acquisition process where the window of opportunity for making efficient MPT decisions is open.

Finally, the iterative nature of the LCSMM dictates that MI MOS-IEW systems analysis must be performed on a recurring basis as the acquisition proceeds. Although pre-Milestone I analysis and decisions are critical, there is a continuing need to review the equipment analyses and decisions throughout the acquisition process because the IEW system itself is changing and being refined. System demands will, in turn, be changing and require refinement. Analysis of MI MOS capabilities and IEW system demands, therefore, must occur iteratively. While the early decisions before Milestone I reflect planning considerations, decisions after Milestone I are basically implementation decisions leading to personnel assignments, training, and IEW system fielding. There is a need for analytical methods which can be used to match MOSs with IEW system demands.

Research sponsored by the Army Research Institute (ARI) is aimed at developing an MI MOS analysis method which can be used by USAICS to address critical personnel and training issues prior to Milestone I. The method being developed is called the "MI Job Comparison and Analysis Tool," or simply "JCAT." JCAT is being designed as a MI MOS analysis method to help answer the following types of questions:

- Can existing MI MOSs provide the skills and abilities required to operate and maintain new IEW systems?

- If not, what are the new demands being placed upon selection, classification, and training MI soldiers in order to operate and maintain these new IEW systems?

Figure 2 presents a conceptual framework for the MI MOS analysis method. There are critical dimensions to the method which can provide the capability to assess the suitability of soldier capabilities to meet IEW system demands. First, there is a need to systematically identify and quantify soldier capabilities in terms of the existing MOS and career management field (CMF) supply. Second, there is a need to identify the IEW system demands in a similar fashion. Third, methods are needed to crosswalk between soldier capabilities and equipment demands determining whether "good" matches exist or not.

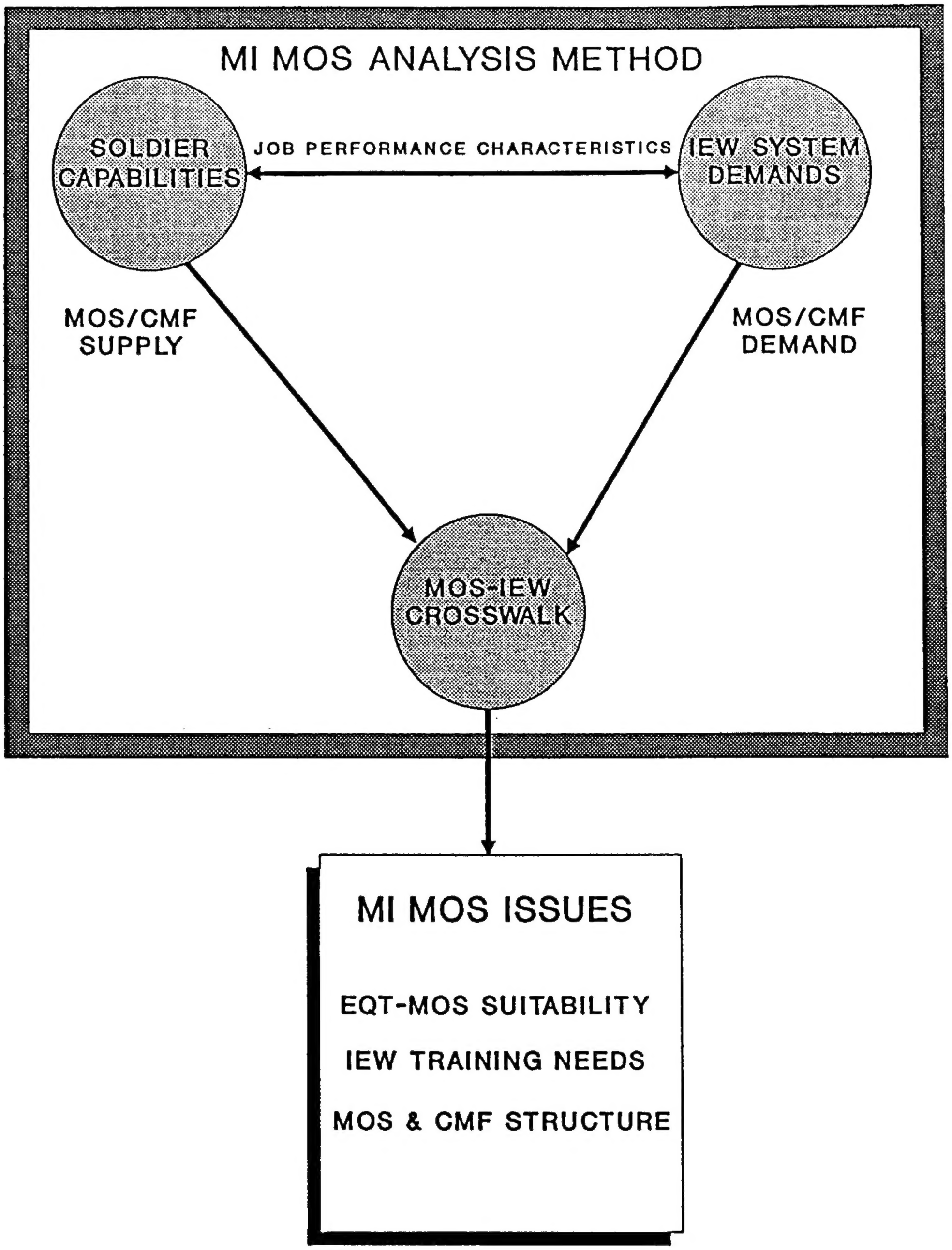


Figure 2. Conceptual framework for MI MOS analytical method

JCAT: A Bridge Between MI MOSs and IEW System Demands

The principal reason for assessing MI MOS capabilities and IEW system demands using JCAT is to determine if existing capabilities can meet equipment demands. If there is a match, then the soldier should be able to maintain and operate the equipment; if there is no match, consideration may be given to different MOS assignments, additional training to enhance the MOS abilities and skills profile, changes in the operating or maintenance requirements of the equipment, or the creation of a new MOS, among other possible initiatives.

JCAT uses a technique which can be used to solicit information from subject matter experts and other appropriate personnel (referred to as "raters") regarding whether or not particular abilities are required and, if so, how much is necessary to perform particular jobs. Abilities refer to the attributes of the individual performing the task. These attributes range from physical capabilities such as vision, hearing, and strength, to intellectual capabilities like reasoning and logical thought capacity. Abilities are relatively stable, enduring traits. Skills refer to the individual's level of proficiency on a specific task.

The technique is comprised of two parts. First, the raters determine the presence or absence of 50 abilities relative to a particular job. The abilities are listed in Table 1. Second, for those abilities rated, the rater determines how much of the ability is required for performing intelligence production activities. Table 2 lists these.

Decision Flow Diagrams

Figure 3 shows the first page of the binary flow diagrams used in JCAT by raters to determine whether a particular ability is required to perform a particular job. The questionnaire is used with an answer sheet in which the rater circles the abilities which are required.

For ease of use, a standard set of symbols has been used to guide the rater through the process. Triangles indicate instructions, particularly, the flow through the charts. Rectangles indicate questions about the job; the questions have been carefully designed to elicit "YES" or "NO" responses about the job being rated. Ellipses contain the name of an ability and a number referring to the order in which the ability is listed in the answer form.

Table 1. JCAT Abilities

COMMUNICATION SKILLS:	1.	Oral Comprehension
	2.	Written Comprehension
	3.	Oral Expression
	4.	Written Expression
CONCEPTUAL SKILLS:	5.	Memorization
	6.	Problem Sensitivity
	7.	Originality
	8.	Fluency of Ideas
	9.	Flexibility of Closure (Pattern Recognition)
	10.	Selective Attention
	11.	Spatial Orientation
	12.	Visualization
REASONING SKILLS:	13.	Inductive Reasoning
	14.	Category Flexibility
	15.	Deductive Reasoning
	16.	Information Ordering
	17.	Mathematical Reasoning
	18.	Number Facility
SPEED-LOADED SKILLS:	19.	Time Sharing
	20.	Speed of Closure
	21.	Perceptual Speed and Accuracy
	22.	Reaction Time
	23.	Choice Reaction Time
PERCEPTUAL SKILLS: VISION	24.	Near Vision
	25.	Far Vision
	26.	Night Vision
	27.	Visual Color Discrimination
	28.	Peripheral Vision
	29.	Depth Perception
	30.	Glare Sensitivity
PERCEPTUAL SKILLS: AUDITION	31.	General Hearing
	32.	Auditory Attention
	33.	Sound Localization
PSYCHOMOTOR SKILLS:	34.	Control Precision
	35.	Rate Control
	36.	Wrist Finger Speed
	37.	Finger Dexterity
	38.	Manual Dexterity
	39.	Arm Hand Steadiness
	40.	Multi-Limb Coordination
GROSS MOTOR SKILLS:	41.	Extent Flexibility
	42.	Dynamic Flexibility
	43.	Speed of Limb Movement
	44.	Gross Body Equilibrium
	45.	Gross Body Coordination
	46.	Static Strength
	47.	Explosive Strength
	48.	Dynamic Strength
	49.	Trunk Strength
	50.	Stamina

Table 2. MI Production Activities

Planning: Any intelligence processing activity or group of activities which involves preparing in advance how you intend to accomplish a task or job function. For example, outlining a set of questions to ask a subject, determining how equipment must be deployed, determining the frequencies to collect on.

Setting-Up or Preparing: Any intelligence processing activity or group of activities which must be accomplished before a mission related task can be carried out. For example, deploying equipment, calibrating equipment, collecting information from a data base, preparing a map overlay.

Collecting Data: Any intelligence processing activity or group of activities which must be carried out in the collection of data which will later be processed or analyzed by you or someone else. For example, interrogating a subject, listening to and recording voice communications, watching signals on a scope, operating collection equipment.

Managing or Cataloging Data: Any intelligence processing activity or group of activities which prepares the collected data for later processing or analysis. For example, using a computer terminal to input data, logging the receipt of a spot report in a journal, placing incoming information on a sitmap.

Analyzing or Exploiting Data: Any intelligence processing activity or group of activities which requires the processing of collected data or information; to combine it into a higher level of information or to determine the relationships between various types of information. For example, doing intelligence preparation of the battlefield, providing input to an intelligence estimate.

Interpreting Data: Any intelligence processing activity or group of activities which result in a prediction from or an explanation of a body of previously analyzed data. For example, deriving possible avenues of approach, developing alternative deception strategies, figuring out why the enemy has made an unusual move.

Preparing Outputs: Any intelligence processing activity or group of activities which requires data or information to be placed in a format. For example, making briefing charts, putting data in message format, encrypting.

Disseminating Information: Any intelligence processing activity or group of activities which result in the transmission of information or data from one source to another. For example, sending a message by morse code, delivering a briefing, talking on the radio or telephone.

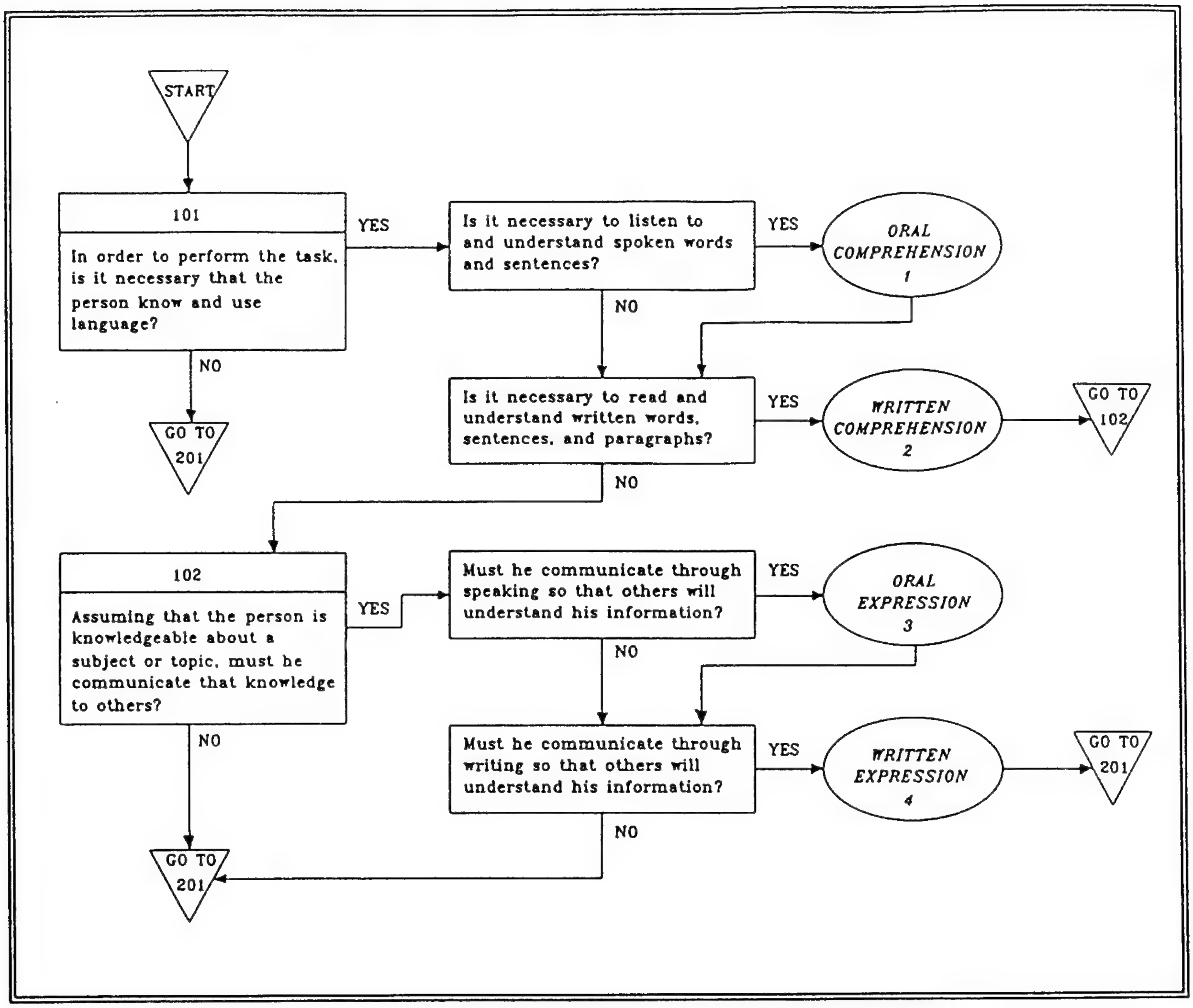


Figure 3. JCAT decision flow diagrams

If the rater reaches an ellipse through the decision flow process, the ability contained therein has been judged as required to perform the job being rated. The rater would interrupt the decision flow process to record the answer on the form provided. This process would continue until a determination has been made for each of the 50 abilities.

Application to Intelligence Production Activities

The second part has been tailored to elicit information specific to Military Intelligence. The rater is asked to indicate how much of the abilities selected in the first part are required in each of the eight intelligence production activities. Where there is a requirement, the rater uses a seven point scale to make the estimate. To facilitate making the judgements, anchors based on "great amount," "quite a bit," "moderate," and "minimum amount" of the ability are provided. Figure 4 provides a sample of the JCAT form used to elicit and record data pertaining to the intelligence production tasks.

Raters applying JCAT to determine MOS abilities and skills profiles generally work independently once the survey administrator has explained the procedures and methods. Approximately 1-2 hours are required to use JCAT to assess the abilities associated with a MI MOS. While applications of JCAT to new IEW systems have not yet occurred, a more involved, iterative group process could be required. Unlike MOS rating where it is assumed a common job experience exists, a common view of a new IEW system, particularly in the early stages of an acquisition, often does not exist. The additional time required for JCAT applications to IEW systems is needed to develop a common definition of the job performance requirements that will exist for the new equipment.

For the abilities you selected, how much is required for .....? (Place your estimate of the value in the appropriate cell.)

- 7 A GREAT AMOUNT OF THIS ABILITY IS NEEDED
- 6
- 5 QUITE A BIT OF THIS ABILITY IS NEEDED
- 4
- 3 A MODERATE AMOUNT OF THIS ABILITY IS NEEDED
- 2
- 1 A MINIMUM AMOUNT OF THIS ABILITY IS NEEDED

	PLANNING	SETTING UP OR PREPARING	COLLECTING DATA	MANAGING OR CATALOGING DATA	ANALYZING OR EXPLOITING DATA	INTERPRETING DATA	PREPARING OUTPUTS	DISSEMINATING INFORMATION
39. **ARM-HAND STEADINESS.** The ability to keep the hand and arm steady. It includes steadiness while making an arm movement as well as while holding the arm and hand in one position. This ability does not involve strength or speed.								
40. **MULTI-LIMB COORDINATION.** The ability to coordinate movements of two or more limbs (for example, two legs, or one leg and one arm), such as in moving equipment controls. Two or more limbs are in motion while the individual is sitting, standing, or lying down.								
41. **EXTENT FLEXIBILITY.** The ability to bend, stretch, twist, or reach out with the body, arms or legs.								
42. **DYNAMIC FLEXIBILITY.** The ability to bend, stretch, twist, or reach out with the body, arms, and/or legs, both quickly and repeatedly.								
43. **SPEED OF LIMB MOVEMENT.** Involves the speed with which a single movement of the arms or legs can be made and/or repeated. This ability does not include accuracy, careful control, or coordination of movement.								

Figure 4. Example of JCAT answer form

Potential Uses of JCAT Data

The goal of collecting MOS abilities and skills data is not simply to analyze those abilities and skills but to be able to provide a crosswalk between the information about MOSs and IEW system demands to address major personnel and training issues faced by USAICS. As an indication of the potential use of JCAT, consider the following areas.

MANPRINT Requirements and Opportunities

Within the framework of the LCSMM, the Army has initiated the MANPRINT program to enhance achievement of performance goals by considering soldier capabilities and limitations as integral elements of total system performance during the acquisition of hardware and operational software systems (AR 602-2). The policies and procedures stemming from MANPRINT require analyses such as discussed here and create a framework in which analysis using JCAT can be accomplished.

USAICS, as the IEW combat developer, is responsible for initiating MANPRINT prior to Milestone 0. The combat developer performs or coordinates early studies, analyses, and evaluations on the proposed IEW system to determine initial MANPRINT requirements. The combat developer establishes and chairs the MANPRINT Joint Working Group (MJWG) which is a tailored organization responsible for determining the level of MANPRINT involvement in the system development and plan all MANPRINT inputs and activities for the entire system life cycle.

USAICS, as the training developer and the personnel proponent, is responsible for performing early analyses on the proposed IEW system to determine training requirements, the need for training devices, and training constraints imposed by the new system. In its role as personnel proponent, USAICS is responsible for determining the personnel supportability implications of the new system, including accessions and personnel life cycle.

The membership of the MJWG is determined by the proponent and is based upon program needs and the nature of the acquisition. Its composition may be altered as the acquisition progresses. In response to the interdisciplinary nature of MANPRINT, the working group's membership usually includes representation of proponent members with vested interest including the combat developer, the training developer, the personnel proponent, and, the materiel developer, among others. Each of these members could contribute to JCAT analysis.

New System Design

In the course of developing new IEW systems, tentative system designs are proposed. These have to be evaluated from many perspectives. One dimension is in terms of the demands for abilities and skills. Using the current JCAT, an abilities and skills profile can be developed for the new IEW system. The profile is diagnostic in that it gives suggestions as to where to look in the system design for the tasks creating the demands.

Another use of these data is to compare alternative new designs or as a tool in so-called "comparability analysis" required in the HARDMAN and MANPRINT methodologies evaluating a new design against a predecessor. The method here is to assess differentially the abilities and skills needed to perform tasks in the old and new systems or the alternative new systems.

Training Requirements

JCAT data can be used in establishing new training requirements by determining the abilities and skills being trained versus the IEW system demands. Also possible is a kind of "training comparability analysis" evaluating established training against new training requirements. The general question concerns the degree to which training must be changed as jobs change.

One particular issue of some importance is identifying general abilities and skills requirements and then training for these general skills prior to specific system training. How exactly effective this might be remains an empirical question.

Finally, many of the new MANPRINT methods require extensive quantitative tradeoff studies between training variables and other manpower and personnel dimensions. One may, for example, ask questions about the optimal balance of selection and training variables in achieving high levels of skill performance. JCAT abilities and skills data are ideal for this type of tradeoff.

Manpower and Personnel Actions

Every day in organizations decisions must be made about critical manpower and personnel actions. As one example, the anticipated introduction of a new system or set of jobs may require manpower not available in a fixed-level force. Actions have to be taken on such questions as "Which member of the existing force are most applicable to the new system and what will be the human resources consequences of their transfer?" Another example is the requirement to decrease the quantity of the force, and questions arise such as "What will be the impact on the abilities and skills of the manpower pool with anticipated reductions?"

Beyond quantity will be questions of quality, "How adequate are present personnel to perform new jobs and what additional training, if any, will have to be provided?" JCAT data are particularly important to the question of personnel quality demands. Abilities and skills are identified, and the level of demand is both a qualitative and quantitative indication of personnel quality requirements. Conversely, the aptitude levels of the soldiers available will effect performance and these data can be used to predict anticipated aptitude level requirements.

Selection Criteria

It is a well established truism that the closer selection parameters are to job dimensions the better the selection of personnel will be (many other things being equal). One problem in developing selection criteria is to have sufficient detailed information about the tasks, jobs, and job families and for that information to be in a form to be used in developing and using selection tools. JCAT data in abilities and skills are particularly appropriate for selection development.

Selection for the armed forces is a massive undertaking both conceptually and operationally. Since World War I, the United States has done a remarkable job in creating and sustaining an outstanding selection system. Constant efforts are being made to improve the system such as the current ARI Project A which, in part, is increasing the validity of selection tests and composites by extending the use of the Armed Services Vocational Aptitude Battery (ASVAB). But a major and ever-continuing need will always be to generate detailed job and task data, such as anticipated abilities and skills sets, to which selection criteria can be matched.

MOS and CMF Analysis

A very complex structure has been provided for Army personnel with respect to jobs and job families in the MOS and CMF system. But the requirements for MOSs and CMFs are under constant change, and methods are needed for rational and effective transitions that will maximize effective personnel classification and utilization and minimize personnel hardships.

JCAT abilities and skills data provide a psychological basis for understanding MOSs and CMFs. The data can be used as one basis for evaluating the currency of MOSs and CMFs. The data can assist in changing MOSs and transitions among MOSs. As shown earlier in this chapter, the data can serve to evaluate the fit between existing MOSs and future system requirements. While not sufficient alone for MOS and CMF analysis, abilities and skills data can play a very effective role in MOS and CMF analysis and the subsequent actions taken to improve career management.

Working Paper

WP HUA 91-02

Intelligence Production as a Model for Processing Information

David D. Burnstein, ARI Field Unit-Ft. Huachuca

August 1991

Reviewed by: Julie A. Hopson
JULIE A. HOPSON
Chief, ARI Field
Unit-Ft. Huachuca

Approved by: Julie A. Hopson
JULIE A. HOPSON
Chief, ARI Field
Unit-Ft. Huachuca

Cleared by:__________________
ROBIN L. KEESEE
Director, Systems Research
Laboratory

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

INTELLIGENCE PRODUCTION AS A MODEL FOR PROCESSING INFORMATION

CONTENTS

List of Tables

List of Figures

INTELLIGENCE PRODUCTION AS A MODEL FOR PROCESSING INFORMATION

Abstract

Most behavioral research in information processing has focused on how individuals process information rather than organizational environments where interconnections of people and machines process information. To adequately address issues arising within large and complex information processing systems, it is necessary to broaden the individual focus. This paper describes an effort to expand the individual focus to one with an organizational perspective. A conceptual model of organizational information processing, developed for military intelligence production, is described and the use of the conceptual model to determine relevant research questions and approaches is discussed.

Introduction

Human information processing is a complex area of study. It exists at many levels, from processes occurring in the sensory cell, through the central nervous system control of cognitive operations, to how organizations process information. Most behavioral research has focused on how individuals process information rather than organizational environments where interconnections of people and machines process information. This focus is evident in research on military intelligence (MI). Although military intelligence is a large and complex information management and processing system, research has primarily focused on soldier functions.

Implicit in the individual approach is that effective prediction, diagnosis, and modification of soldier performance, will benefit the system. This may not be the case. In military intelligence, both individuals and machines contribute to the system output. As a result, a change to any part of the system must be considered in relation to its effect on the entire system. More importantly, changes cannot degrade the system product. Thus the effective prediction, diagnoses, and modification of performance within the system must occur in terms of relationships occurring within the system and be based on standards compatible with the requirements for the system's performance.

While individuals and organizations use information in producing their products, MI is an example of a pure information processing organization. MI requires both raw data and processed information as its raw material; its basic functions are information processing functions, and its output is intelligence, i.e., information, both descriptive and predictive. MI represents

a large and complex system in that it requires the integration of a relatively large number of people, often widely dispersed geographically, working at many different organizational levels, using a variety of equipment of different complexity, to produce the various tailored outputs. Table 1 is a generalized list of the characteristics of the MI system.

The Intelligence Production Conceptual Model

At the lowest level of specificity, characteristics of the MI system described in Table 1 can be conceptualized as a simple input-process-output model (Figure 1). However, the intelligence production system supports all levels of command, is comprised of well defined hierarchical organizational elements which must interact in order to produce intelligence, and the intelligence production tasks occur to some degree in all functions within each element at each level of command. Therefore to be useful, the conceptual model must depict the complexities of intelligence production at a higher level of specificity than the simple input-process-output model.

At another level, the conceptual model can be represented as a network. In the conceptual model shown in Figure 2, the nodes (circles) represent functions required to produce intelligence. Though not shown, the intelligence production tasks are nested within the nodes. The links (lines with arrows) between the nodes represent where the product of the node is next used. The product of the nodes is represented by the $\triangle$ (delta). The use of the $\triangle$ is to emphasize the changing nature of the information as it passes through the production system. The lower case letters on the links help to describe the paths information flows as it is transformed by the MI production system.

Structurally and functionally, the conceptual model should include the intelligence production characteristics described in Table 1. Structurally:

1. It retains the input-process-output characteristics.

2. More than one function (node), is necessary to produce intelligence.

3. Information/data can flow from one function to another, as depicted by line ac (link), or from one function to several functions, as depicted by links be and bd.

4. Information/data can be received by one function, as indicated by link ab, or received from multiple functions, as indicated by links ef and cf.

Table 1

Characteristics of the MI system

Relationships With the Users
1. It supports a hierarchial Army command structure by providing intelligence (processed information) both descriptive and predictive.
2. It is structured in a hierarchy which supports the equivalent level of the command structure.
3. The output of each level of the MI hierarchy is tailored to the needs of the command structure at that level.
Input/Output Considerations
4. It uses information (processed and unprocessed) as its raw material and produces processed information as its final output.
5. Different levels of the MI hierarchy can receive common information or information unique to that level of the hierarchy.
6. Within the hierarchy of the MI structure either unprocessed information (raw data) or processed information is passed to the next higher level.
7. Within the hierarchy of the MI structure only intelligence or combat information is passed to the next lower level in the hierarchy.
Processing Considerations
8. At any level within the MI structure, intelligence production (the processing of either raw data or processed information) can be described in terms of the production functions required to produce the intelligence.
9. Without respect to the level of the MI hierarchy or functions, each function may require the same information production tasks be carried out, although not necessarily to the same degree.
10. These common information production tasks tend to be: planning, collecting, managing, analyzing, integrating, interpreting, preparing, and disseminating information.
11. The information production is accomplished by humans, machines, or a combination depending on the structural level of the MI system and the production functions being carried out.

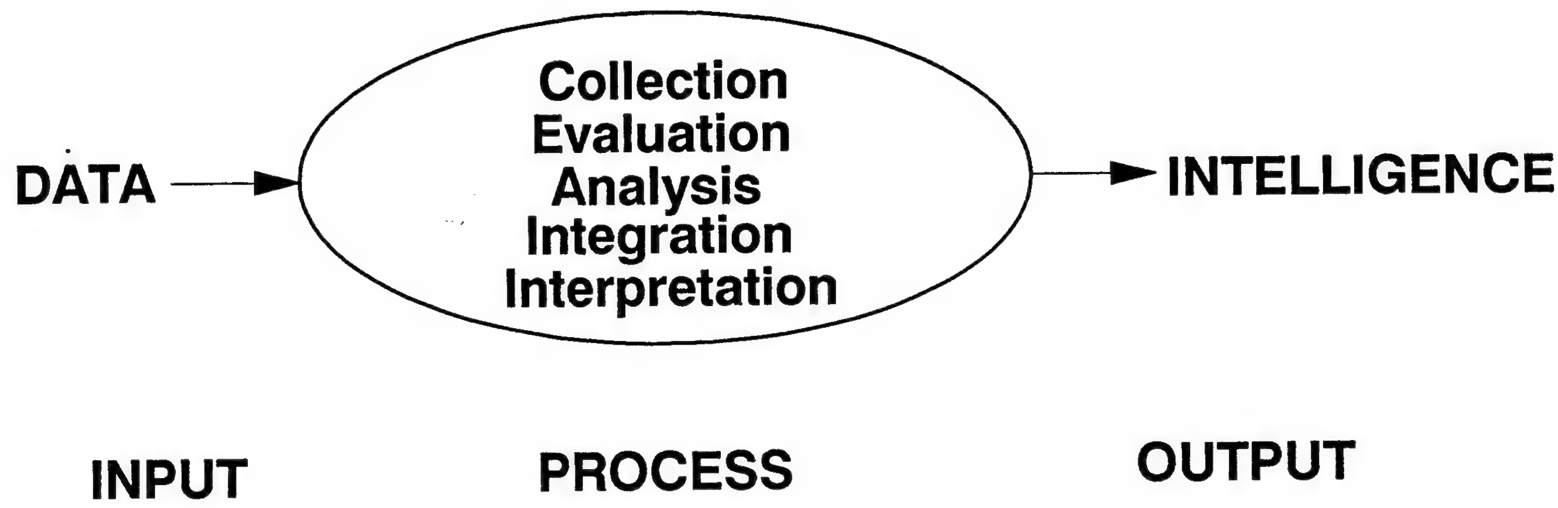


Figure 1. Intelligence production as a simple input-process-output model.

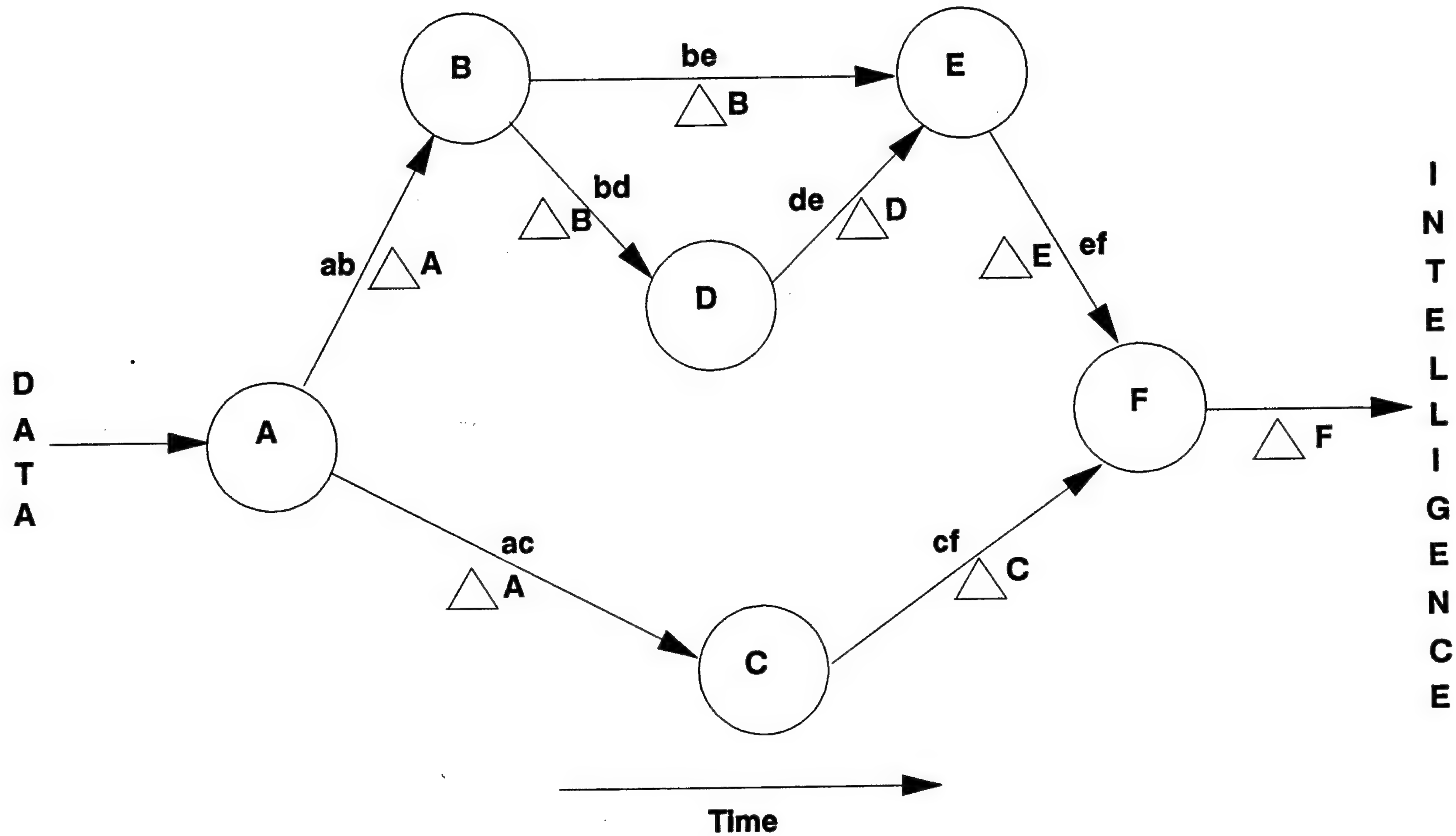


Figure 2. The IEW intelligence production conceptual model.

5. The transformation of data to intelligence can follow a simple path, for example, links ac, cf, or a more complex path, for example, links ab, bd, de, ef.

6. The transformation of data to intelligence occurs in one direction, as indicated by the arrows on the links, and over time, as indicated by the directional arrow labeled time.

Functionally

7. The nodes contain the intelligence production tasks required to change the input to output (from one △ to another).

8. Each node has a specific output, identified by a lettered △ .

9. The nodal △ are inputs to other functions. Thus, node B acts to transform the output from node A (△A) to its own output, △ B, which is sent to nodes D and E. △F is the final output from the system.

10. Since intelligence production occurs within the node, the output, △, must represent the results of production processes or tasks.

Figure 2, can be regarded as a conceptualization of one level of the MI system. In terms of the level of command structure supported, the conceptual model can be expanded for descriptive purposes (see Figure 3). The arrows between echelons only indicate that information flows between various levels of echelon. The figure does not represent the actual interrelationships between the functions at different levels.

A more important expansion of the model is the decomposition of the functional nodes. Without respect to a particular function or level of command within the intelligence production hierarchy, each functional node can be decomposed to the level of specificity necessary to address a particular problem. Figure 4 represents the decomposition of a function to a generic intelligence production task level and identifies dimensions for describing the task. These are:

1. The information requirements are the data requirements necessary for the task to be carried out. Included in the data requirements is their source. The source helps to identify the relationships between functions at different levels in intelligence production hierarchy.

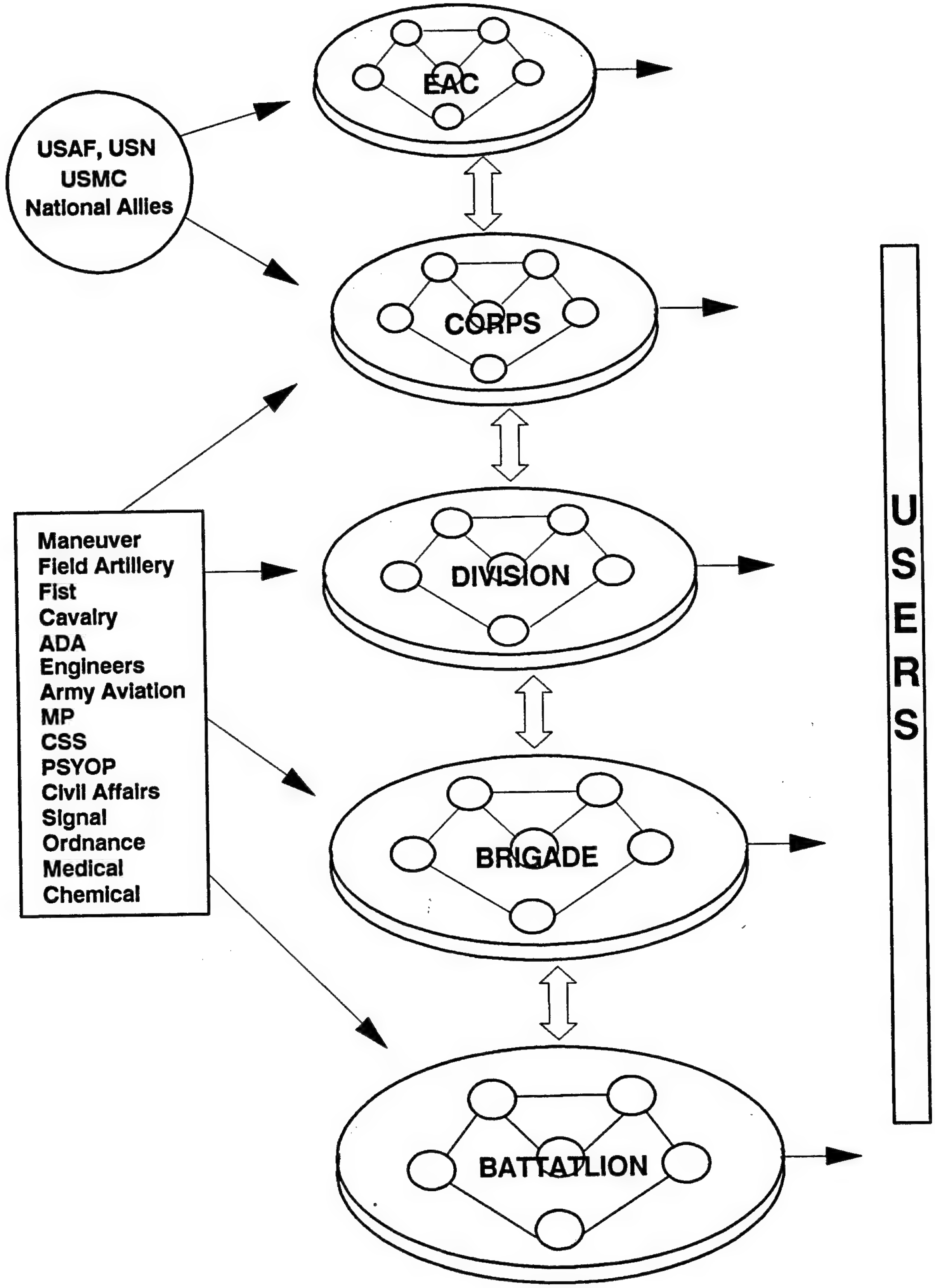


Figure 3. An expansion of the conceptual model through echelons.

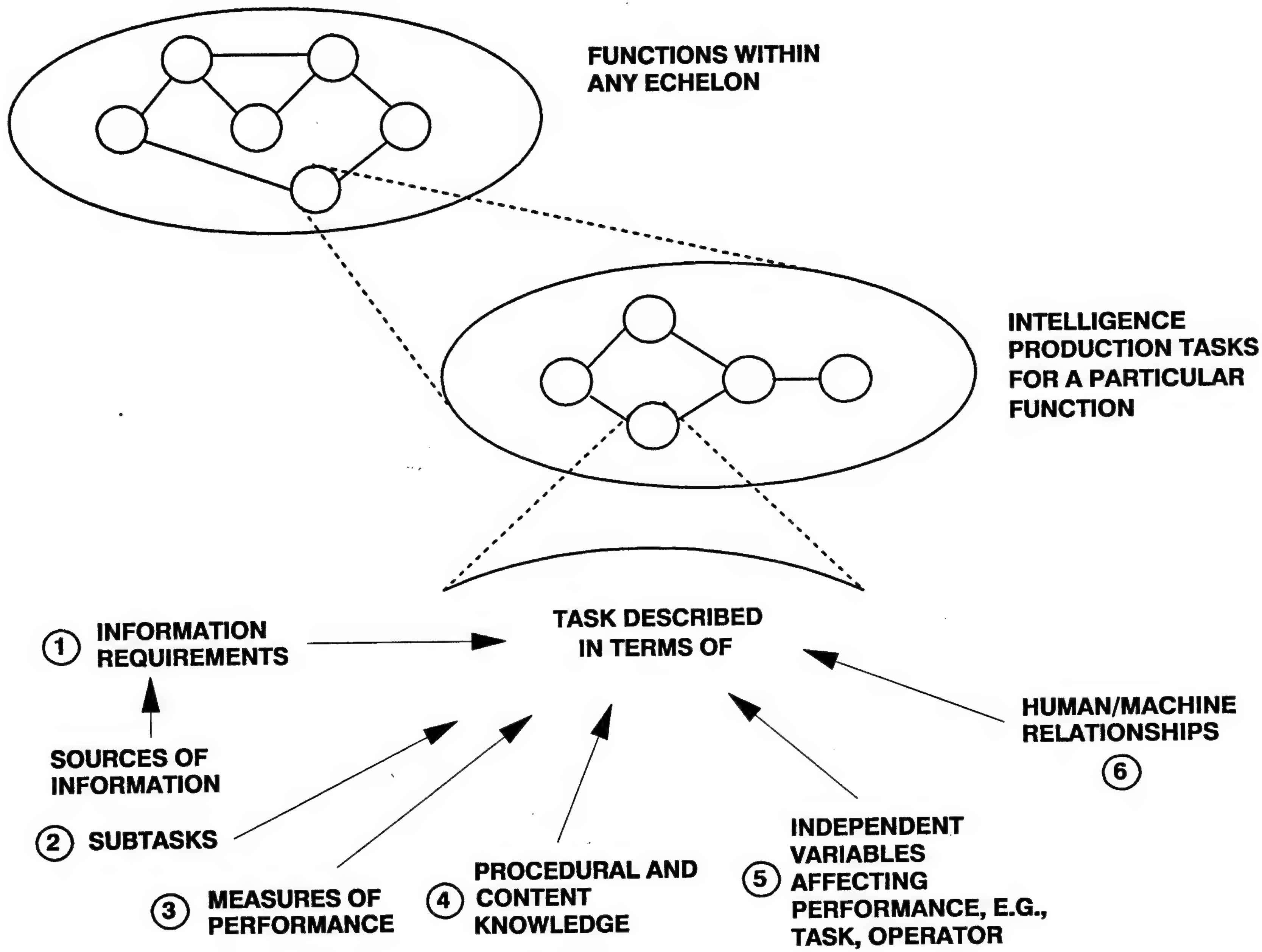


Figure 4. Decomposition of the conceptual model to a production task level.

2. The sub-tasks which make up the task are included because they often help to identify and clarify some of the other elements of the task description. In addition, they set the boundary for the most specific level of decomposition.

3. One of the critical purposes of the decomposition is to identify the relevant dependent variables to measure. Both the measures of effectiveness and measures of performance are required.

4. Procedural and content knowledge specify the knowledge required to carry out the task.

5. The independent variables affect behavior and influence the output. They may be derived from the task (e.g., level of difficulty), the environment in which the task is carried out (e.g., workload), or from the operator (e.g., skill level).

6. Human machine relations can be in terms of what is currently used or may be used in the future.

The level of task description, like the level of decomposition of the function, depends on the problem being addressed.

Measuring Intelligence Production

In order to effectively predict, diagnose and modify the intelligence production of both individuals and the organization, relevant measures are required. Although the conceptual model is descriptive, it provides a framework for identifying what, where and how to measure the elements of the system.

Implicit in the conceptual model are three types of measurement: measures of performance (MOP), measures of efficiency (MOI), and measures of effectiveness (MOE).

MOEs are measured against the standards required to perform a function, be it a node within the MI system or an external user. For example, in Figure 2, the effectiveness of output $\triangle$ F is intelligence user defined, while the effectiveness of $\triangle$ A is defined by the requirements of nodes B and C. The nodal MOEs are also dependent on the system MOEs. Even though the standards for the MOEs are defined by the requirements of nodes which use the output, nodal MOEs must be compatible with the system MOEs. An appropriate decomposition of the intelligence production conceptual model can facilitate ensuring the compatibility of the MOEs by depicting the appropriate backward chaining for MOE development. As a result, the model serves as a kind of control

to ensure the final intelligence product is acceptable to the user rather than reflecting what MI "wants" to give the user. MOEs are expressed in terms of quantity and quality.

MOPs are task dependent, that is they are measured within a node. They measure the behavior required to produce an output. There can be many different measures of MOPs depending on the level of specificity of the behavior. Most MOPs within MI can be measured in terms of latency, time to perform, or number of errors.

MOIs are measures of efficiency. They are made within the nodes and represent the cost of changes in behavior or effectiveness. Measures include:

Time costs--that it takes a longer or shorter period of time to carry out the later functions,

Manpower costs--the time to carry out the function remains the same, but it now takes more or less people to perform the function,

Error modification--it forces or eliminates errors from occurring within later functions,

Transmission errors--passes errors onto the next function that were not characteristic before the change,

Opportunity costs--the gain or loss of time and manpower at one function may have positive or negative consequences for unrelated functions, and

Psychological costs--changes could also result in, for example, increased stress or frustration which can be measured independent of their impact on effectiveness or resource costs.

The three measures permit the impact of change in the intelligence production system to be determined either internally within a node, at various nodal outputs at different levels of the intelligence production hierarchy, and externally in terms of the system's performance.

Any change is expected to enhance the effectiveness or efficiency of intelligence production, and stems from trying to remedy a system dysfunction or evolutionary enhancement of the system. The impact of the change is measured as value added. Value added is a relative concept which can have either a positive or negative value. It requires the comparison of the measurements resulting from the change to be compared to the measurements before the change.

In reference to Figure 2, value added can occur within a node. This would be accomplished by making changes in the production tasks or how the tasks are carried out, for example, automating a task. The value added could be measured either in terms of efficiency in carrying out the task, or in terms improvement in the output (effectiveness). Value can also be added by changing the input to any node, for example, inputting pre-processed information rather than raw data. Again, either efficiency or effectiveness can be measured for value added. Finally, value could be added by changing the path of information as it flows through the production process. In addition, value added can be determined within nodes or at outputs not directly effected by change. For example, a change to the input $\triangle$ A could be measured in terms of the value added to the output $\triangle$ F. Figure 5 summarizes the measures of value added.

The Model as a Tool

The purpose of the conceptual model is to have a high level descriptive framework which can be used to guide and structure questions concerning prediction, diagnosis and modification of performance within the intelligence production system. The conceptual model should help to examine the various levels of specificity inherent in intelligence production, identify the relevant MOE and MOP, and determine where the measures need to be taken. It should help to determine how to identify areas of deficiency and assess possible remedies.

Diagnosing the Intelligence Production System

The conceptual model provides a framework for diagnosing deficiencies in intelligence production. It implies that successful intelligence performance is a result of the adequate functioning of the entire intelligence production system. With reference to Figure 2, this means that the impact of any dysfunction within the production system will be manifested by deficiencies within $\triangle$ F. If the MOEs have the appropriate dimensions and sensitive scales, then the deficiencies in the $\triangle$ provide clues to the system dysfunctions and their locations.

The model implies that the criteria for measuring effectiveness for any $\triangle$ is determined by the perceiving node. Thus, the node that must process that input determines the criteria the input must meet. For example, the user of the intelligence would determine the criteria for $\triangle$ F and the criteria for $\triangle$ A would be determined by its users, nodes B and C. In addition, the model indicates that only processes within a node produce the $\triangle$. Therefore, the causes for deficiencies manifested by the $\triangle$, lie within the nodes. Since the model indicates that $\triangle$ F, the final performance, is the result of a

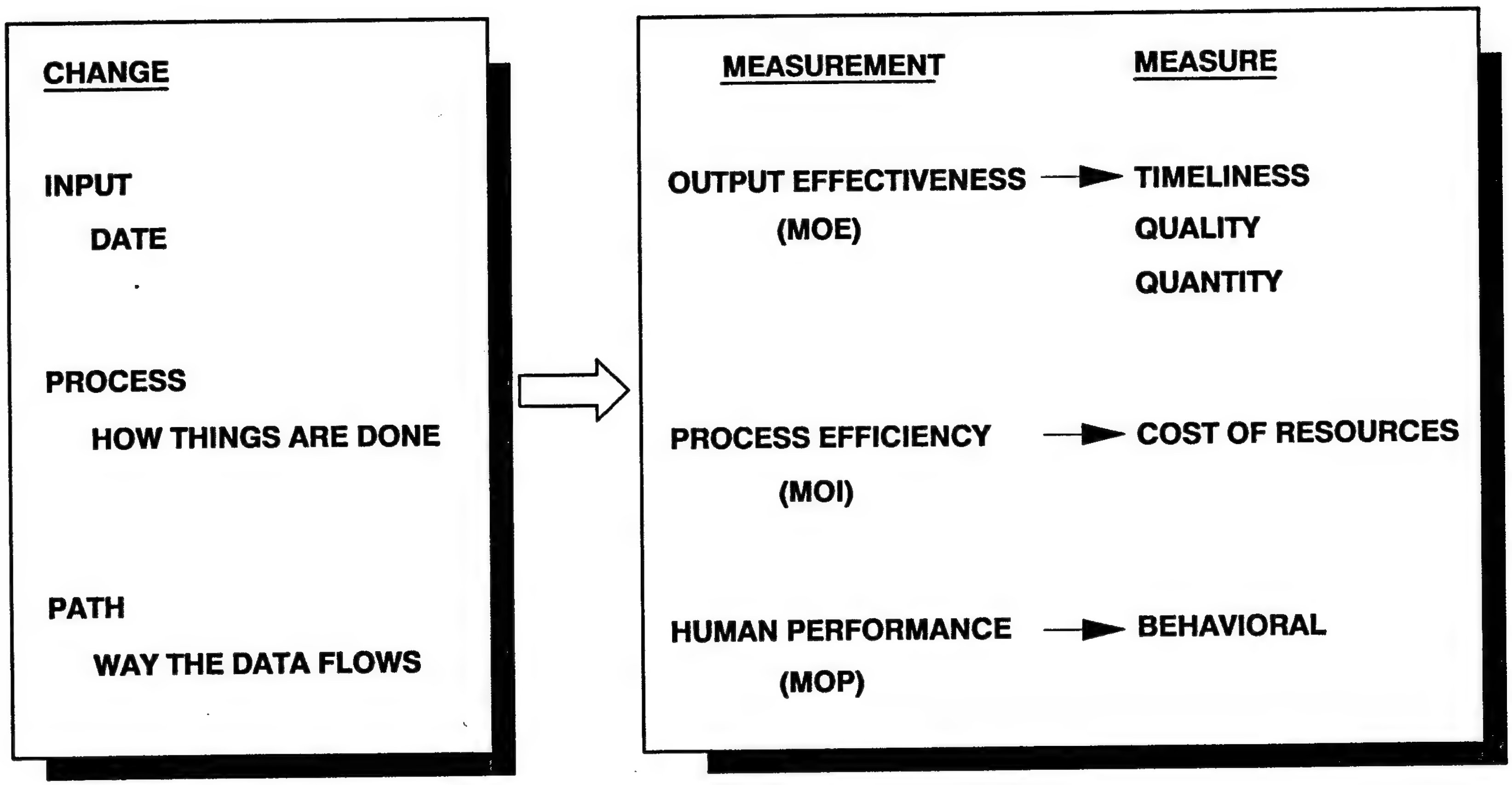


Figure 5. Measures of value added.

sequential operation, the implication is that diagnosis should be done in the reverse sequence. Thus, if $\triangle$ F was deficient, the implied diagnostic strategy is to diagnose node F for dysfunction. However, the first diagnostic step would be to determine if the input $\triangle$, E and C, met the criteria standards imposed by node F. If they did, then it would be appropriate to diagnose node F. However, if either or both $\triangle$,E and C were deficient, the next diagnostic step would be to diagnose the nodes that produced them. In diagnosing node E and F, the first diagnostic step would be to determine if the input $\triangle$ met the standards for the nodes.

The overall diagnostic strategy is to proceed backward through the production sequence until the point that there are no deficiencies in the input $\triangle$ for the node having an output deficiency. The diagnosis for identifying a dysfunction would begin at that node. The same diagnostic strategy would hold within the node. The inputs to production tasks would be analyzed first to determine if there were any deficiencies, then tasks having satisfactory input, but deficient output, would be diagnosed for deficiencies.

The diagnostic, as outlined, is dependent upon being able to specify and measure the output of the functions within the intelligence production system. It requires as its beginning point a scale on which to identify the standards for the input $\triangle$.

<u>Designing Modifications</u>

Change to the intelligence production system is required in order to maintain or enhance current performance. The direction for change comes from many sources, from lessons learned during war to the imposition of personal preferences. The directed modifications can appear in the form of changes to doctrine, force structure, training, and materiel. Since what drives the changes is expectation of enhanced performance, the model could help guide the design of the change.

The model implies that any change within the intelligence production system will influence the entire system. As a result, the model can help the designer of the modification to visualize how the change will reverberate through the system. This should indicate what additional changes might be necessary and cue the possibility of unwanted impacts.

Since the model requires MOE for output, a performance state for the system and its functions can be defined. In turn, modifications can be defined in terms of what and how much performance should be expected to change. As a result, the designers of the modifications have the system's performance criteria to use to guide the design.

In reference to Figure 2, an example is if a modification was proposed for function D. By using the model, several questions become readily apparent:

Given the current performance, $\triangle$ D, what changes must be made in the function D to increase the new $\triangle$ D to the new criteria?

If the modifications are made, will changes have to be made to $\triangle$ B for the modification to be effective? If yes, what does the change in input mean to function B?

Will the function of E change as a result of the new output $\triangle$ D? If no, will changes in the processes within function E have to be changed to accommodate the new input?

If the modification in function D imposes a modification on function E, what is the impact on the performance of function E?

The questions can go on ad nausea, but the point is that the model can be used as a framework for asking the questions. Furthermore because the model requires MOE and MOI, there is the capability of predicting how much and what impact the modification will have throughout the entire intelligence production system.

<u>Measuring the Impact of Change</u>

The model can also be used as a framework for assessing changes made to the system. Changes include actions taken to remedy a system dysfunction or the implementation of planned modifications for enhancing system efficiency or effectiveness. The changes can be assessed either in an operational or modeling environment.

According to the conceptual model, change can be implemented at a node, at an input $\triangle$, in a path, or some combination. A new SOP, the addition of a material system, or decrease in manpower would be examples of changes implemented within a node. Increasing the amount or kind of information that must be processed by a node would be examples of changes to input $\triangle$. With reference to Figure 2, an example of change in the path would be sending $\triangle$ D directly to node F rather than node E.

The effect of any of these changes is measured at the output $\triangle$ of the affected node(s). In the above example, the effect of changing the path would be measured at node F. As previously discussed, the measure of the effect of the change is value added or removed. If a new data processor was being used (a change within a node), value added would be the difference between

output using the new processor and output not using it. The performance would be measured in terms of effectiveness and efficiency.

Since the model implies that successful intelligence performance is the result of the adequate functioning of the entire intelligence production system, the effect of any change should be measured at all the subsequent $\triangle$ in the path. Thus a change to node B could result in increased value added to $\triangle$ B, but the change in $\triangle$ B as input to nodes E and D could result in decrease in value added of outputs $\triangle$ D and E. The model implies how and where to look for unplanned repercussions of change.

Model Application

The model provided a measurement and diagnostic framework during a contract for developing a methodology for evaluating Military Intelligence Unit Effectiveness in an operational setting. Research had identified the information requirements of the intelligence users and a procedure for setting the priorities for the requirements. With reference to Figure 2, this was $\triangle$F. However, since the model called for dimensions and standards, the priority ratings were insufficient as a measure of effectiveness. As a result, five dimensions, with scales, for measuring effectiveness evolved. The dimensions were timeliness, frequency, operational perspective, clarity, and completeness.

In addition, the dimension of timeliness requires a referent. It was only with reference to the model that we realized that timeliness had several different referents, and that the one required for our methodology had to be operational rather than production. The model helped to establish the dimensions and made us aware that while the dimensions of the internal $\triangle$ and the output $\triangle$ might be the same, they have different interpretations.

It had been established that a fault analysis would be used to determine the causes for deficiencies within the intelligence production system. However, it was still necessary to have a strategy for where to go in the intelligence infrastructure to look for causes. The model provided the framework. The strategy developed was based on backward engineering through the model using structural/functional nodes of low resolution.

Summary and Conclusions

A descriptive model of the intelligence production system was presented along with the implications of the model for diagnosing production deficiencies, designing system changes and evaluating implemented change. Based on one case, it appears the model can provide a productive framework for developing ideas.

Working Paper HUA 88-03

MEASUREMENT AND EVALUATION OF MILITARY INTELLIGENCE (MI)
PERFORMANCE: SUPPLEMENTAL REPORT

Theodore C. Fichtl and John R. Thompson
Science Applications International Corporation

Contract Monitored by Jo Hall

June 3, 1988

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

MEASURING THE PERFORMANCE OF MILITARY INTELLIGENCE

Is there a tool that the United States Army can use to measure the effectiveness of Military Intelligence (MI) performance? Are users of intelligence satisfied with the intelligence they are receiving? Both are enormously important questions that are addressed in this paper.

Is There a System for Measuring MI Performance?

There can be little debate that the MI unit Skill Qualification Test (SQT) and Army Test and Evaluation Plan (ARTEP) only measure individual and unit task/skill competency. They do nothing to measure the effectiveness of the total intelligence infrastructure from sensor operator in the MI Battalion to the G2 and the tactical operations support elements. Other branches of the Army have methods and quantitative measures of effectiveness that permit end-to-end effectiveness evaluations. For example, consider the world of the Field Artillery. An artillery battery's function is to destroy identified targets. Effectiveness of the battery is a direct measure of the target value and how close the rounds come to the target.

Is there a corollary within MI of an 'identified target' and, perhaps more importantly, what determines how close the MI 'round' comes to hitting the target? Additionally, who identifies the intelligence target and the value of the target to the requestor? If MI's 'rounds' are the myriad of outputs resulting from exercising the intelligence process, then what are the measures used that define 'on target '

performance? There are currently no answers to these questions that represent a rigorous, reliable evaluation system.

Why Isn't There an Evaluation System?

A significant part of the problem may be the absence of a useful definition of what the 'target' of intelligence really is. An additional component of the problem may be the ambiguities that have historically surrounded the question of the needs of the users of intelligence. Intelligence doctrinal publications should provide the required 'target' definition. Unfortunately, a finite useful definition cannot be found. What is reflected in doctrine is vague and highly generalized. Specifics on 'hitting the user target' with intelligence are often expressed in terms that emphasize product type and reporting cycle. There is little useful reference to information that will actually satisfy user needs.

Is this merely the result of contemporary oversight or are there other contributing factors? Actually, the origins of the situation are found in military history dating back to the Revolutionary War. For example, General George Washington provided guidance on the 'intelligence target' in December 1776 when he wrote to Brigadier General Maxwell,

> You are to be extremely vigilant and watchful to guard against surprises and to use every means in your power to obtain a knowledge of the enemy's numbers, situation, and designs . . . Every piece of intelligence which you may think of importance for me to know, communicate it without loss of time.

Washington's guidance is somewhat vague and generalized. It is difficult to deduce what may have been important and what may not have been beyond "numbers, situation, and designs." Historically this type of guidance, which almost suggests mind reading, unfortunately has persisted. However, in 1936 after WWI, Major Edwin E. Schwein, an instructor at the Command and General Staff College, attempted to bring rigor and clarity to the world of intelligence by publishing Combat Intelligence, Its Acquisition and Transmission. In his forward, Schwein states,

> Since the WAR, probably no other section of the general staff has been so much a mystery to the average officer as the second, or intelligence section..We all know that the principal role of the commander is to make decisions . . . Obviously then, information of the enemy forms the base for all intelligent decisions.

Schwein then develops a systematic approach for ensuring that collection, interpretation and dissemination of enemy information provides that base. Unfortunately, he abandons developing the idea of 'information of the enemy' per se and focuses on form. Schwein observed,

> Our service has adopted three convenient forms for the dissemination of this intelligence. They are:
>
> a. The summary
> b. The intelligence report
> c. The intelligence estimate.

We still view intelligence in terms of these three artifacts. In fact they have shown

no substantial change since 1936 even though the nature and pace of ground combat has changed. To this day, a strong implication remains that satisfying the requirements for form in some way will ensure that intelligence output will inherently be useful.

What About Today's Doctrine?

Current MI doctrine continues to place emphasis on form. Only reporting schedule has been added as a rigorous criteria for intelligence output. For example, both FM 34-1, Division Intelligence and Electronic Warfare Operations, and FM 34-3, Intelligence Analysis, stipulate the form of intelligence. They further point out that most intelligence reports are simply published in accordance with an established unit schedule. The guidance on formats and cyclic production is designed to ensure that potentially useful information is included in any given output. This subjective guidance is illustrated by the following extract from Chapter 3, FM 34-1 that describes an Intelligence Summary (INTSUM):

> The INTSUM contains a brief summary of information of intelligence interest covering a period of time designated by the commander. The INTSUM provides a summary of the enemy situation in the forward and rear areas, enemy operations and current situation and updates other intelligence reports. Negative information may be included in the INTSUM, but unnecessary information is excluded. The INTSUM reflects interpretations and conclusions of enemy capabilities and probable course of action. It has no prescribed format except that INTSUM will be the first item of the report. However, when involved in joint service operations, originators of INTSUMs will use the format contained in Chapter V, JCS Publication 12. Nonessential detail should be excluded from the INTSUM, but information concerning the issuing unit, DTG of issue, brief discussion of capabilities and vulnerabilities, and conclusions should always be included.

The guidance that "negative information may be included in the INTSUM, but unnecessary information is excluded" is interesting, but as it was in 1776 and 1936, in 1988 it provides no objective basis for determining what is "negative" or "unnecessary." The same criticism and confusion results from the guidance that "nonessential detail should be excluded from the INTSUM." Where is the objective guidance to help determine what is nonessential? The natural effect of this ambiguity is to place premiums on production volume rather than usefulness of information. It also tends to foster length -- not conciseness -- in intelligence outputs and promotes the production of intelligence for the sake of intelligence. An unfortunate conclusion is that the intelligence system is conditioned to create output that assumes qualities of uniformity and conformity rather than usefulness.

OK, There May Not Be Much in Doctrine, But What Do Users Say?

Users of intelligence should be able to provide insight into how well Intelligence is 'on target.'

In 1987, under the sponsorship of the Army Research Institute (ARI), a study entitled "Measurement and Evaluation of Military Intelligence Unit Information Processing Performance" was undertaken. As part of the effort, an Intelligence Product Utility Questionnaire was circulated to individuals representing the intelligence user community including G3s, Fire Support Coordinators (FSCs), and others. Responses to the survey illustrate some interesting and unexpected

views on the intelligence producer/user relationship. The following are extracts of the findings derived from the questionnaires.

1. How well does doctrine define useful intelligence products?
Finding. In many cases the doctrinal definitions are highly generalized. Many of the numerous outputs of intelligence are found to be useful by users, though not all useful products are doctrinally defined. Missing are many of the more current outputs of the intelligence process such as Intelligence Preparation of the Battlefield (IPB) results and other informal products. Guidance that defines user requirements for intelligence is generally absent.

2. What are the outputs of MI?
Finding. In the responses, outputs were defined in several ways: in terms of doctrinal product definitions; in terms of interactions between intelligence and users that would include informal briefings, graphics, databases and results of IPB; or in terms of relevant information items.

3. Does one intelligence product fit all users?
Finding. Probably not. The survey results showed distinct differences in perceived information needs, the relative importance of information to what users do, and a preference for outputs, not simply products, tailored to what users need.

4. Do intelligence outputs differ in terms of their usefulness to users?
Finding. Yes, though there is a range of difference across types of users. The preliminary results indicate that the more useful products are high in their contextual nature (i.e., when information is in terms of a combined enemy/friendly/area/time context). Products ranked lower in usefulness tend to be information without context in terms of age, detail, or association with only one aspect of the situation.

5. Who are the users?
Finding. Responses tend to indicate users should be grouped by the SIGMA STAR structure (Maneuver, Fires, Air Defense, Intelligence/Electronic Warfare, and Combat Service Support). The original view of users was by FM 101-5 definition of command staff organization and interactions. FM 101-5 is regarded as obsolete in view of emerging Airland Battle doctrine.

6. Is intelligence itself a user of MI outputs?
Finding. Many G2s have difficulty viewing themselves as a user of intelligence. This response appears driven by conventional views of intelligence output as those doctrinally identified products.

7. Are all users equally important?
Finding. The data are inconsistent as to user importance. Doctrine does not specify relative importance of intelligence users but individuals questioned had differing opinions. The G2 tends to consider operations (G3, G3OPS, G3PLANS)

as the dominant users. Specific questioning of G3s, however, failed to verify that this bias was perceived by operations.

8. Who is responsible for defining user needs/requirements?
Finding. While the user is typically viewed as ultimately responsible for defining information needs/requirements, it is not always reasonable for intelligence to assume the user can accurately state them. Users often don't know what to ask for from intelligence because of the highly technical nature of collection capabilities and not knowing what might be specifically important.

9. Do users have the same information needs?
Finding. Probably not. Although there was consistency in general categories of information needs, individual users had distinctly different ratings on specific information item importance. The responses suggested that specific information items take on added importance as a function of a situation.

10. Does the user's set of specific tasks impact information needs?
Finding. No, information needs seem to be independent of the task being performed. There are influences on a user's information needs, but 'task at hand' does not appear to be a major one of those influences.

11. Is there a difference between wartime and peacetime information requirements?
Finding. The information item requirement list is basically the same between peacetime and wartime, but the relative importance changes from item to item from peace to war.

12. Are there situational factors that affect the use of intelligence?
Finding. Yes. Situational considerations are very important in the use of intelligence and should generally be influenced by five factors: mission, level of conflict, echelon, who the user is, and the application for the information.

13. What is the current view of the user community toward intelligence outputs?
Finding. There is a general user perception that MI products are not as good as they could be. Additionally, MI seems to share a negative view of its own current products.

14. Is information overload a real problem?
Finding. Information overload appears to be a problem related to failures of intelligence to filter irrelevant information and to deliver information in a usable form. There is no indication that information overload occurs when information is delivered in situation display form or through briefings with accompanying graphics.

15. Is information underload a real problem?
Finding. Yes. Certain users apparently live with a dearth of relevant MI output.

16. Can users perform better with alternative intelligence output forms?
Finding. Probably yes, if for no other reason than they would be more likely to use the information formulated 'on the fly' which reflects concentration on the user and situational dimensions.

17. Are users of intelligence actively seeking and developing alternatives to doctrinally defined and formatted products?
Finding. Yes, a graphic INTSUM has been developed and is being experimented with in place of the lengthy textual version at several divisions.

18. What is the role of IPB in the production user oriented intelligence?
Finding. IPB is emerging as an extremely important process -- both in defining the nature of effective intelligence outputs and also in defining distinct levels of information in basic weather and terrain templates, as well as situation and event templates. IPB apparently defines a process structure for MI that may directly reflect on the type of interactions with users and the types of outputs generated. Several types of IPB outputs are not comprehensively correlated as doctrinal products. These include overlays and templates of enemy, terrain, weather and situation, event and decision templates .

19. Can the utility perspective of the user of intelligence become part of the intelligence process?
Finding. There appears to be emerging acceptance of a concept for measuring the user utility of intelligence information that may permit matching the nature of intelligence outputs to user information requirements, even within a diverse range of situational possibilities.

The results of the ongoing field research indicate there is a need to develop a method for evaluating the effectiveness of the entire intelligence process. The research also indicates, however, that the development of a measurement instrument for MI Unit Effectiveness can not rely exclusively on the guidelines provided by MI doctrine. Comments from the field interviews indicate that to measure the distance intelligence outputs are from the 'target' requires the users of intelligence to be brought into the picture. The survey results are representative of opinions that are at odds with current doctrine and practices in many cases. The survey participants, all professional officers, offer new views on the general level of effectiveness of MI and lend impetus to the development of an MI effectiveness measurement tool that can enhance performance. The concept of analyzing the degree to which

intelligence outputs are utilitarian to the users provides the framework for developing the required tool.

User Utility and Measuring Military Intelligence Performance

A research hypothesis is that MI performance can be measured as a function of the degree to which intelligence outputs, the products of the intelligence process, meet user utility requirements. First, however, outputs must be characterized in a way that permits evaluation. The characterization can be based only on product forms, or for more rigorous analysis, can be based on variables that typify intelligence output in terms of content. The latter approach is more consistent with parallel research being conducted by the Army in such areas as Commander's Information Needs. Secondly, based on the indications found in the survey data, other aspects, such as situational variables, must be defined and their effects addressed. Finally, users must be identified in a fashion that permits validation of differing views of utility.

Research Overview

The ARI research effort entitled "Measurement and Evaluation of Military Intelligence (MI) Unit Information Processing Performance" has produced the prelimary results outlined above. This research has a series of tasks illustrated in Figure 1 that will yield a methodology for evaluating the performance of MI in terms of user utility and diagnostics for tracing faults and identifying remedies.

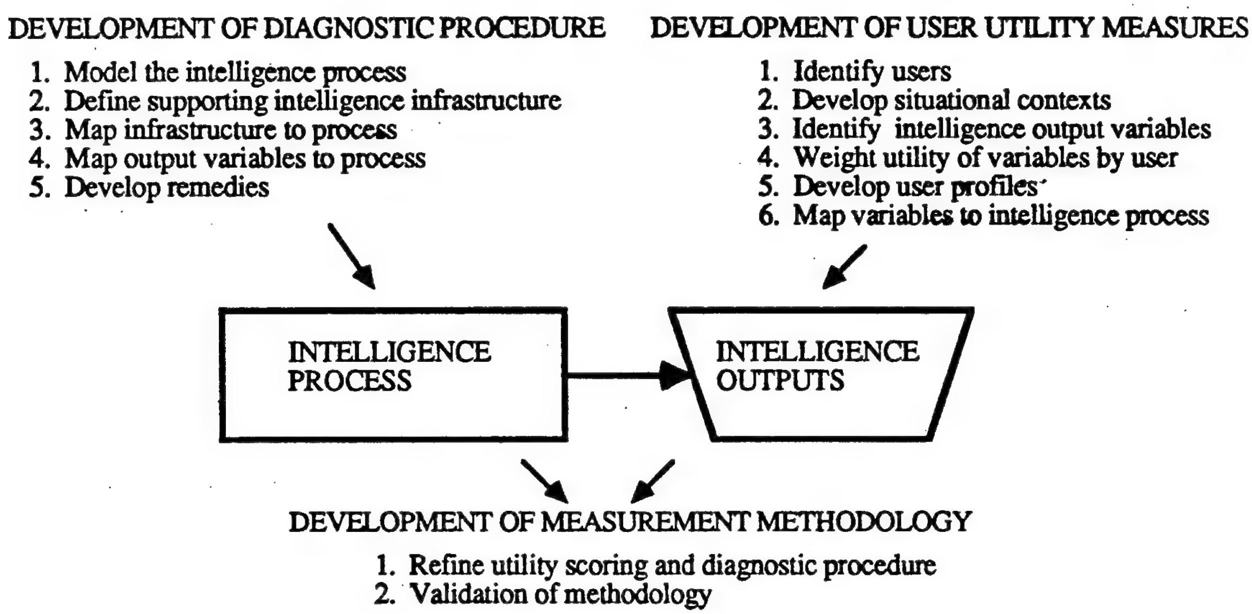


Figure 1. Research tasks to develop a measurement methodology

Research is under way in six major areas of investigation that represent the basis for the MI effectiveness measurement methodology: (1) defining intelligence outputs in terms of relevant information to users; (2) creating profiles of user information needs in different situational contexts; (3) defining critical elements of intelligence output utility that can be scored by a user ; (4) creating a diagnostic procedure to trace a low user utility score to a responsible element of the MI infrastructure; (5) identifying responsible processes and organizational elements of the MI infrastructure for given types of low utility scores and relating faults to

remedies; and (6) refining the scoring and diagnostic procedure for use in a training or operational environment. Additionally, this research addresses the problem of creating prescriptive guidelines and standards for training.

Research Objectives

The development of a tool to measure the utility of intelligence output meets one essential goal, that of providing the MI community a rigorous evaluation methodology. An additional benefit will be the ability to diagnose why intelligence outputs fail to meet user utility requirements and where and how the failures are induced. Further, the measurement methodology will permit the creation of prescriptive guidelines in the form of user profiles to provide a model of user information needs for the training environment of a means of establishing effective intelligence outputs in operational environments.

When completed, the measurement methodology as illustrated in Figure 2 will assist in the process of providing high utility, user-oriented output, while permitting the producers to exercise quality control in tailoring outputs to well defined user needs.

ARI's research is intended for application in both the training and operational environments. If the methodology is ultimately adapted to software, it will offer an opportunity to optimize the formulation of intelligence output for the user for the first time. In the automated classroom environment, the methodology

offers the opportunity of training and evaluating student performance from the perspective of the user.

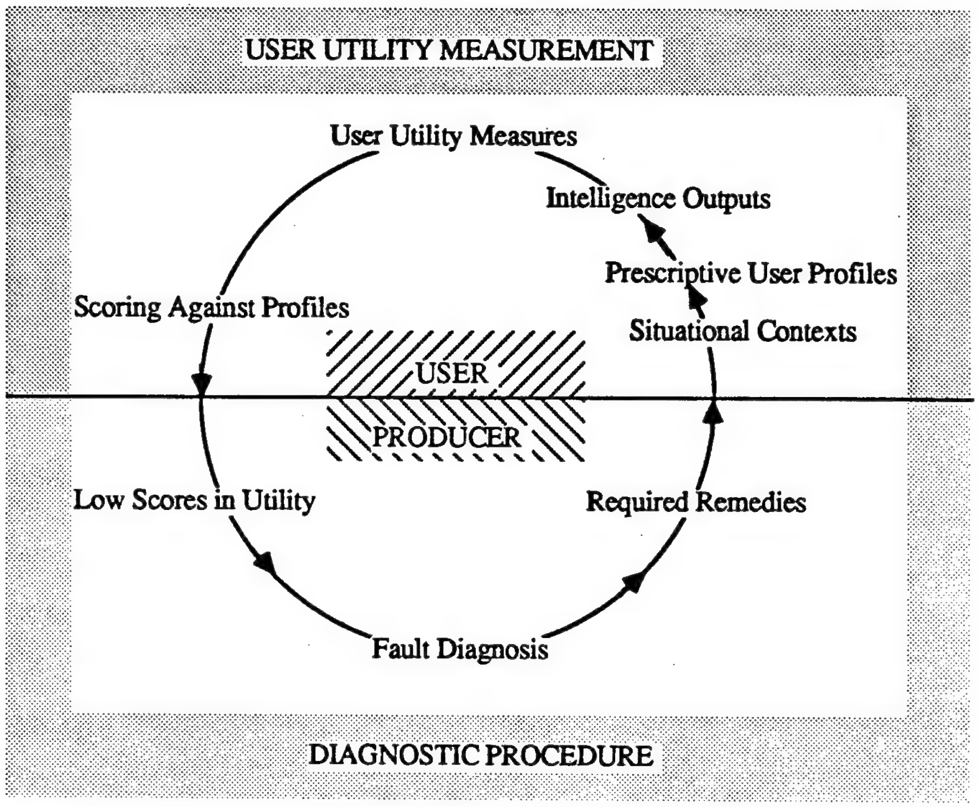


Figure 2. Military Intelligence effectiveness measurement methodology.

Working Paper

WP HUA 89-03

The Role of the Human Operator in the Morse Code Collection Cycle: A Descriptive Model

Beverly G. Knapp, ARI Field Unit-Ft. Huachuca

September 1989

Reviewed by: JULIE A. HOPSON
Chief, ARI Field Unit
Ft. Huachuca

Approved by: JULIE A. HOPSON
Chief, ARI Field
Unit Ft. Huachuca

Cleared by: ROBIN L. KEESEE
Director, Systems Research
Laboratory

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

THE ROLE OF THE HUMAN OPERATOR IN THE MORSE CODE COLLECTION CYCLE: A DESCRIPTIVE MODEL

CONTENTS

List of Figures

THE ROLE OF THE HUMAN OPERATOR IN THE MORSE CODE COLLECTION CYCLE: A DESCRIPTIVE MODEL

Introduction

The Military Intelligence (MI) community has shown increasing interest in refining the selection, training, and operational performance of morse code intercept operators. This has become important in a time of constrained budgetary and personnel resources. Although morse training has a long standing history of research and development, a more extensive, systems approach to the morse area is now advocated due to continued high attrition rates and the promise of automation technology for both the training and operational environments.

A task force was convened by the US Army Research Institute (ARI) in late 1988 in order to dedicate intensive behavioral science research support toward morse operator issues. Although the primary thrust of the task force is targeted toward modelling the morse code learning process and discovering individual differences in skill acquisition therein, a secondary endeavor involves documentation of the status and impact of automation technology in the morse workplace. Not only does this have a potential impact on future training strategies and requirements, but a persistent question raised by policy and decision makers from outside MI, concerns the relevance and need for morse operator personnel in today's highly sophisticated sensor environment. At issue is the potential for automation technology to redefine, partially eliminate, or eliminate personnel requirements and training demands.

The objective of this report is to document the results of a number of site visits (conducted at the National Security Agency, US Army Intelligence School-Devens, US Army Intelligence Center and School, and HQ Electronic Security Command during the May-September 1989 time frame) and knowledge elicitation sessions with subject matter experts in the morse collection and training environment. These fact-finding sessions were undertaken to clarify the current role of morse operators in relation to state-of-the-art and emerging automation technology, and have resulted in a descriptive model of the operational environment. A second objective of this report is to indicate the implications of the descriptive model on the concurrent ARI learning model research.

Morse Operational Environment

An understanding of the morse collection, processing, and dissemination environment is gained by the recognition of two key facts: (1) morse code is a viable and important communications means used extensively by many countries throughout today's world

to transfer data; and, (2) the threat environment (which includes morse communications) is target rich: meaning that there are simply more intelligence reportable data sources than there exist potential capabilities to collect and process the data.

Communications Intelligence (COMINT), as an MI discipline, consists of four distinct but interrelated areas: morse intercept, non-morse intercept (teletypes, printers, etc.), voice intercept, and facsimile intercept. In addition, the radio/direction finding (RDF) operator function supports the entire COMINT area. Each of these sub-disciplines is concerned with a variety of target "types" or sources, which are networks of communicating entities. Transmissions between entities may be characterized by intelligence value, according to urgency of reportable information, ranging from "highly critical" to "of routine importance" for further analysis. Figure 1 depicts the multidimensional nature of the COMINT threat and target environment.

COMINT is a single source of intelligence data which, when combined with other sources (photo reconnaissance, human and technical intelligence, etc.) can form a larger intelligence picture. Within the single source realm of COMINT, however, morse collection is a microcosm of numerous detection, identification, and preprocessing activities all structured to derive product reportable information from raw target input data.

Morse Processing Cycle

An individual morse operator is not engaged in copying International Morse Code (IMC) in isolation but is part of a larger cycle of events and activities geared to the "servicing" of target information for intelligence production and dissemination. A typical morse "mission profile" actually contains a series of decision points for the utilization of manual or automation activity. Figure 2 is a descriptive model which shows a typical morse mission flow for target servicing. Each point along the event flow is also tagged as currently automated or manually performed (or both, if options exist). Although the model portrays a serial flow of events, actual operations would consist of cycles where this basic process is repeated continually and in parallel, by multiple operators (both manual and automated), in order to "service targets."

The process depicted in Figure 2 begins with a continuous sampling via automated collection means, of the threat environment. Following this, scanning and searching occurs, based on predetermined intelligence information requirements. This searching process may be "directed" to find a specific

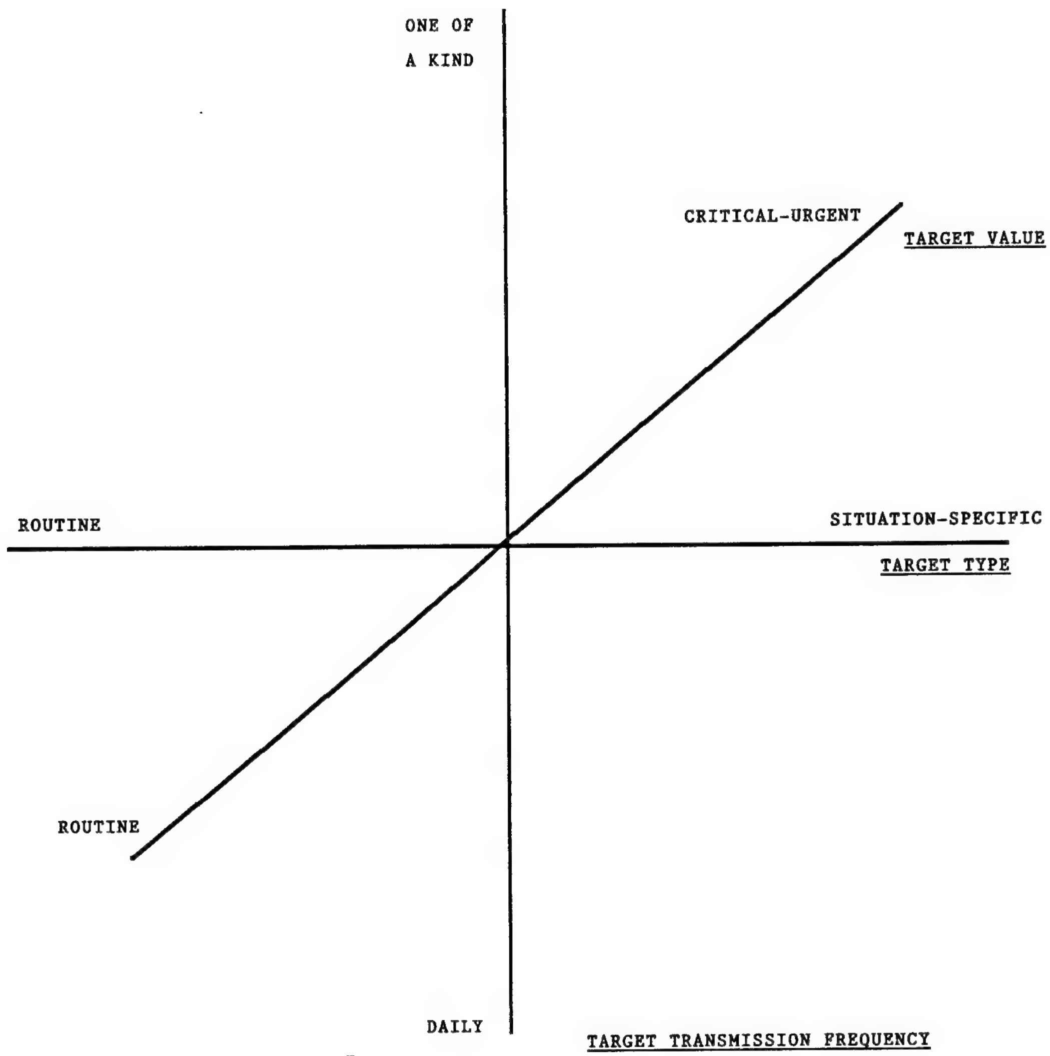


Figure 1. The multidimensional nature of COMINT targets.

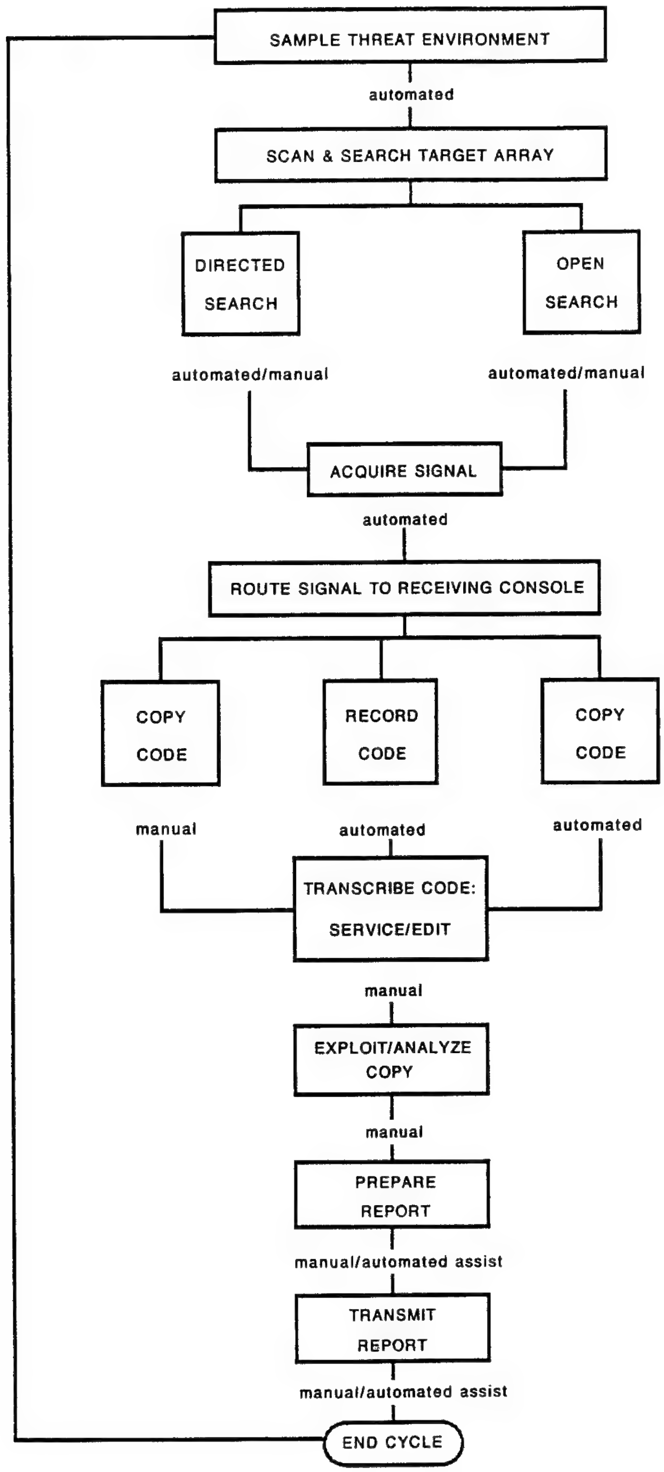


Figure 2. Morse collection event and functional flow.

target on a specific frequency, or "open" to sample signals of possible interest. Either search process may be manual or automated. Certain targets are more easily detected by manual or automated means, depending on equipment capability, operator skill level, clarity of signal (signal to noise ratio), and other target characteristics. The workflow proceeds with signal acquisition. The signal, once selected, is then designated for copy or recording to a worksite console.

The receiving workstation or console is the point at which much attention has been given for fully automating operations. A common assumption is that a morse receiving workstation consists of an operator, at a console, wherein any signal received will be a signal copied, and ready for transmission through designated report channels. Therefore, simply allowing a machine to copy a designated signal and pass it along would seem fairly straightforward. However, this is a simplistic, ideal view, attainable in only the best of circumstances. In reality, the processing of a target signal involves a number of decision and editing processes, in addition to the copy task, and is also a function of factors outside the immediate control of any given individual operator. These factors include such things as operator experience level, equipment sophistication and availability, signal quality, and signal transmission rate. At any given time, the interplay among these types of factors will determine the output processing capacity and efficiency of the entire morse collection center.

Figure 3 shows a number of the critical factors affecting morse signal processing operations, and the levels of effectiveness possible for each factor, as a function of conditions prevailing at any given time. For example, if incoming signals at a given hour or shift are few (low input rate), and transmission speeds of those signals, their quality (signal to noise ratio), and operator availability are also optimal, a high quality return of copy may be realized. In contrast, if highly experienced operators are busily engaging high speed, difficult to copy targets, new targets will be compromised. Available technology will either support or aggravate the collection task. If the consoles, recording devices, and editing options are optimal, then workflow will be facilitated. In reality, overall worksite efficiency will vary widely depending on the value or status of each factor for a given time frame.

If a scalar approach were taken to analyze the operation, each factor level might be assigned a value, from, for example, one to three (1=low impact, 2=moderate impact, 3=high impact), and an "efficiency index" could easily be computed. To index efficiency for a time interval, each factor would be rated, then "efficiency score" could range from a low of six (high efficiency, optimal operations) to twelve (moderate efficiency),

EFFICIENCY RATING 1 - 3	SIGNAL INPUT RATE (Signals per Hour)	TARGET CRITICALITY (Urgency, Intelligence Value)	SPEED OF TRANSMISSION (Groups per Minute)	SIGNAL CLARITY (Signal to Noise Ratio)	OPERATOR AVAILABILITY (Experience vs Inexperience)	AUTOMATION TECHNOLOGY STATUS (Baseline Assist/ Full-up)
1	LO	ROUTINE	LO (Within Manual Copy Range)	GOOD Easy Detectability	HIGH Experienced Available	STATE-OF-THE-ART CONSOLES and Advanced Prototypes
2	MODERATE	MODERATE	MODERATE High Speed Copy	AVERAGE Moderate Detectability	MIXED Experienced Engaged	STATE-OF-THE-ART CONSOLES ONLY
3	HIGH	CRITICAL	HIGH Exceed Manual Copy Capability	POOR Difficult Detectability	LO Experienced & Inexperienced Engaged	MIXED Some Outdated Technology

Figure 3. Factors affecting morse collection processing.

to a high of 18, indicating a very busy operation, stressed along many dimensions. All values in between would indicate levels of efficiency and also provide a means to identify the factors contributing most to a degraded mode.

It is clear that the operator activity in relation to the workflow is a complex interchange between type of target, target activity level, personnel capabilities and resources, and technology. Referring back to the mission workflow diagram of Figure 2, the routing task of signal to console is the job of a mission manager, not an operator. The decision to route the signal is based on an assessment of the factor status levels of Figure 3. If a given worksite does not have any copy consoles with fully automated capability, this eliminates this decision option, and the only alternative in a very busy mode is to record the signal for later manual copy, or interrupt ongoing copy if the new signal is critical in nature. Even automatic copy is subject to operator editing for quality control and insertion of important pre and postamble data. This is shown in Figure 2 as the transcribe/edit function.

An additional function, occurring either simultaneously or subsequent to signal copying, is target exploitation and "analysis." As operators "service targets" with certain regularity, a body of expertise is gained regarding target characteristics, and how these relate to other entities of the intelligence situation. This is a significant function, sometimes referred to as maintaining "net continuity" on assigned targets. When copy "down time" exists, an operator is able to review previous target files, relate recent to older copy, derive trends in transmission schedules, monitor increased or decreased activity, and ultimately make inferences regarding threat activity. This function is supportive to traffic analysts and radio/direction finding operators, and involves coordination and dialogue among these personnel.

A newly assigned operator (one who is a recent graduate of resident morse training at the service school), is intended to "grow" through an ever increasing complexity of decision making capabilities, in order to be of highest value in deriving product reportable target activity, as well as to realize a challenging career position involving much more than simple code copy. Growth to a supervisor or mission manager means additional roles that can remove an individual from a receiving console entirely. These duties include quality control of copy material, interpretation of intelligence requirements directives and reporting requirements, and personnel resource management. All of these functions are currently best suited to human performance as opposed to machine operation. Some automation assistance is available to support databases and detect trends, however, the manipulation of those databases is seen as a largely manual pursuit.

The reporting and transmit report functions are the final stages of the morse mission cycle of Figure 2. Morse copy must be inserted into a predefined report format for transmission to established databases. This is an automation assisted task, in that many consoles contain report "templates" allowing "header" and "trailer" data to be filled in, in addition to the code copy.

Automation Opportunities

At any given time, the morse operator is engaged in one or more of the activities described in the mission workflow above. A morse mission manager will be involved in a signal routing decisions to available receiving consoles which have the capability to appropriately transcribe the signal of interest. Automation provides support to these functions at various levels depending upon the particular worksite. The fully automated collection console is in prototype form in only a few instances. This is largely due to the fact that available Artificial Intelligence (AI) technology (rule-based expert systems to process certain targets) has only recently been available. As with all AI technology of this type, the definition of rules, coding, implementation, and testing are labor intensive and costly. Initial trials have been confined to copy algorithms tailored to targets meeting certain criteria by the worksite. Projections for the full deployment and implementation of this technology span a period of many years.

From the above descriptive model of morse collection and processing, it can be seen that operator functioning extends beyond simple code copy, hour after hour, on an eight hour shift. Automated assistance and fully automated consoles are becoming increasingly available but are by no means offering a replacement for the human element. Since current and future threat projections show an increase in target activities, and the need to relate exploited targets to overall intelligence requirements is critical, the initial impact of increased automation is to enhance morse collection efficiency rather than to immediately eliminate personnel. Manpower savings cannot be expected until sometime in the more distant future when prototype technologies can be deployed more widely and institutionalized across worksites.

Relation to Cognitive Learning Model

The descriptive model of morse mission operations reflects decision points where the introduction of automation technology in the processing cycle can enhance and ultimately redefine operator requirements. It is already apparent that a morse operator's duties extend beyond the copy of IMC. The variety of

functions points out that, although some percentage of duty time is spent on code copy, another (highly variable) percentage is spent on other tasks. Certain targets will always require a highly proficient operator to interpret and skillfully copy; even when resources allow implementation of extensive expert system consoles, many targets will still require diligent servicing by manual means. In the interim, the training of high caliber operators through the service school training base is an important consideration. This is of even greater urgency since attrition in the training base is currently very high (averaging 50%). It is important that training (as well as selection) strategies, be tailored to produce individuals who can cover the spectrum of straight (routine) copy, difficult copy (one-of-a-kind or difficult to detect targets), and exploitation, reporting, and analysis.

The ARI research effort to build and test a cognitive learning model is directed toward an in-depth understanding of the learning process for copying code. Beginning with the processing of individual dits and dahs, the model takes a state-of-the-art information processing approach to isolating the stages of activity in code copy. It then portrays the parallel processing activities that occur in copying work groups at various speeds, in IMC format. Performance curves being developed show where learning and skill acquisition is difficult and subject to bottlenecks, and where, why and how plateaus come to occur in the learning process. This will allow a diagnosis of any shortcomings in the training and selection system so that group and individual differences in performance can be better managed.

The learning performance database will document the variability among individuals and their potential to perform the variety of copy tasks required in the operational sites. As more automation technology is implemented in the workplace, the identified learning capabilities should also relate to the potential to grow into the various other collection functions presented by the diversity of targets. In some cases, an individual less proficient in copying code may still find a role in morse operations, since automated assistance may augment certain skills during the time when experience is being gained on the job. Conversely, highly proficient copiers can be more immediately assigned to cover difficult and demanding target inputs. The performance data gathered with the learning model should permit the identification and categorization of specific skill differences and allow more precise matching of persons to functions.

Summary and Implications

There is no question that the morse signal environment presents viable intelligence collection requirements for morse code operators. Current and projected threat assessments include COMINT inputs, including morse, as product reportable data for intelligence reporting. As automation technology matures and is increasingly implemented in the operational workplace, it is important to model the mission event and functional flows so that human resources are appropriately used to contribute effectively to the overall mission cycle. In addition, as decision options for copying and transcribing increase, it will become more important to ensure that training and selection strategies reflect the operational need. Current modelling efforts are dedicated to the morse learning skill only. Future efforts in modelling the changing operational workplace (in a more detailed and quantitative fashion than the descriptive model outlined above) would allow a better link between the learning model, and would achieve a high degree of accuracy in determining human-machine functional allocation in the operational environment. In the face of demographic trends that predict a diminishing source of qualified morse operator candidates, the optimization of human functions is a timely endeavor which should ultimately result in manpower cost savings.

Although this effort is limited to examining the performance of morse operators, the area of COMINT, as discussed earlier, includes three other functional areas (non-morse, voice, radio/direction finding). These three functions are also undergoing changes in human-machine functional allocation due to advancing technology and sophistication in the target environment. Since there is some similarity and overlap in certain COMINT tasks (e.g., net analysis, distinguishing signal from noise, recognizing transmission patterns, cataloging and reviewing target files, etc.) the potential to realize manpower and cost savings in certain common functions exists. Again, a modelling approach to an overall mission cycle would serve to ensure that, as roles become redefined by technological advances, the training, selection, and assignment strategies for manpower resources are appropriate to the operational need and available personnel inventories.

Systems Research Laboratory - Ft. Huachuca FU (Alexandria)

Working Paper

HUA 88-01

Human Factors and Human Performance Issues in Softcopy Image Analysis

George W. Lawton

December 1987

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

Human Factors and Human Performance Issues in Softcopy Image Analysis

George W. Lawton
U. S. Army Research Institute
Alexandria, VA

FOREWORD

The Ft. Huachuca Field Unit of the Army Research Institute is currently conducting research on manpower, pesonnel, training and performance issues in tactical intelligence systems. The availability of imagery in electronic form may have implications for the battlefield of the future. New systems for processing, analysis, and dissemination of imagery for tactical use will have to be developed and fielded. The research described in this paper is directed toward understanding the capabilities of the individual operator in digital imagery processing and analysis. Such an operator will be at the heart of new tactical imagery systems.

EXECUTIVE SUMMARY

Research Objective. This research was done to identify human factors issues which must be considered in the development of systems for the exploitation and dissemination of imagery in softcopy form for tactical use. The factors identified will be integrated into a substantial research project to develop quantitative models of the performance of imagery analysts doing softcopy imagery exploitation tasks.

Research Methodology. Three distinct methods were used in this research. First, the literature on human factors of Video Display Terminal was surveyed to determine if there were any problems which would indicate insurmountable difficulities in the use of digital imagery analysis with Video Display Units or Graphics Display Monitors. Second, Army imagery analysts with experience in both hardcopy and softcopy exploitation gave us their opinions on the subject. Third, we surveyed the state of the art in imagery analysis workstations, both hardware and software, from several fields, including medicine and geophysical analysis.

Research Results. A review of the existing literature on the use of VDUs in the workplace showed that there is no conclusive evidence that they pose any sort of a safety or health hazard. There is some indication that they may aggravate the visual and attentional decrements which naturally accompany close visual work. Interviews with imagery analysts show that they prefer to do hardcopy exploitation rather than softcopy exploitation. Several hypotheses were developed as possible explanations of this result based on follow-up interviews with the analysts and their supervisors. Based on existing softcopy image exploitation workstations in the fields of medicine and geophysical analysis, we believe it is possible to rapidly develop a prototype of such a workstation for research and development purposes. In the paper, we list the hardware and software capabilities such a prototype workstation should have.

Future Research. Based on the literature reviews, product reviews, and the interview results, we propose a large-scale experimental project, using sequential research methods to identify the relevant variables in image analysis performance and to develop a quantitative response-surface model relating the variables of speed and accuracy in image analysis.

OVERVIEW

Problem

The U. S. Army is currently developing the capability to collect and process imagery ingelligence in electronic form. New systems which might be developed and fielded could collect imagery in or convert imagery to electronic form which can be processed in digital form using electronic workstations. The use of digital image processing and analysis is almost completely unknown at the tactical level, and raises a variety of issues concerning hardware, personnel, manpower, training, and human factors. In this paper, we are concerned with issues raised by performance of individual operators using digital image processing workstations.

The performance of individual image analysts doing digital image processing at an electronic workstation for purposes of tactical intelligence production is almost unexplored. In the following paper, we ar going to summarize the results of our preliminary investigation of the factors which effect this performance, with emphasis on the design of workstations and work environments. We will then briefly describe plans for future research to investigate operator performance in more detail and to develop the implications of this research for system design, manpower, personnel, and training.

Research Objective

The objective of the current research is to review and summarize the existing knowledge relevant to the design of workstations and work environments for production of tactical imagery intelligence from digital imagery. In this research, we have been especially concerned with those factors which may have influences on the performance of individual image analysts.

Methodology

To accomplish the research objective, we have reviewed existing literature, and interviewed a group of knowledgeable people on the subject. Our literture review was based on existing reviews (National Research Council, 1983; Helander et al., 1984) and a search of the literature from 1985 to the present. Interviews were conducted with 6 96Ds and 2 of their supervisors during a large field exercise. These 96Ds were selected due to special circumstances in their assignment which allowed them the choice of hardcopy or softcopy image processing.

PROBLEMS IN SOFTCOPY IMAGE PROCESSING

Review of Literature on VDU Utilization in the Workplace

The change from hardcopy imagery to softcopy imagery is in some ways parallel to the changes taking place in civilian office jobs with the introduction of Video Display Units (VDUs) to the workplace. VDUs are being used in jobs ranging from clerical and data entry, through secretarial typing and editing and on up the pay scale to so-called "creative" jobs in management, advertising, scientific research, and engineering. A lot of attention is currently being given to the impact of such workstations in the workplace by human factors researchers and ergonomicists. This topic is receiving special attention due to pressure from labor unions in both the United States and Europe. Labor unions, and other concerned parties have raised questions about the potential harm which may be done as a consequence of the use of VDUs. They are especially concerned about a high frequency of complaints from workers about "eyestrain", and "visual fatigue" due to the use of VDUs on the job. This type of issue should be of special concern to designers of materiel for use on or near the battlefield. The working conditions are far from optimal, and it is very likely that the work load would be very high. We decided that it was especially important to examine the evidence for degraded performance and detrimental effects of VDU utilization and to weigh the importance of this evidence for design of workstations for tactical digital imagery analysis.

The reader should note that the literature reviewed below applies primarily to conventional office-type VDU displays and not to the specialized workstations used in image analysis. The results of this review and certain existing standards for VDU workstations are included here since they provide parameters within which prototype workstations can be designed. But there is considerably more research and development to be done in this area.

To evaluate effects of VDUs on workers in the civilian work place, the National Research Council of the National Academy of Sciences conducted a large scale study. This study evaluated a wide variety of evidence for visual fatigue and other problems related to the use of VDUs on the job. In summary form, the main conclusions of this study are as follows:

- Complaints and symptoms of job related ocular discomfort, musculoskeletal discomfort and stress are common in users of VDUs. Evidence suggests that this is at least as much due to the way the VDU changes the structure of their job as it is to the use of the VDU itself. People in lower level jobs have far more complaints than people in so-called "creative" jobs. Data-entry

clerks have an especially high incidence of such complaints, yet they spend most of their time focusing their attention on source documents and rarely look at the VDU except to verify their accuracy.

- Managers frequently fail to apply well-established principles of good design and practice to jobs when the VDU is introduced. There are a number of principles and standards which should be used to guide selection of image display characteristics, workstation layout and furniture design, illumination, and task design (IBM, 1984; Kroemer, 1983; Snyder, 1983). But they are not widely applied.

- "Visual fatigue" and "eyestrain" are not meaningful terms in any medical sense. The National Research Council recommended that these terms be replaced by more descriptive terms such as "ocular discomfort" and "changes in oculomotor function."

- The evidence that the visual sysmptoms resulting from use of VDUs are different from those resulting from the performance of other near-visual tasks is equivocal. Further research on this topic is required.

- There is no conclusive evidence that VDU use causes permanent damage to the visual system. But the National Research Council cautioned that this is an open question. Below we will note that there is some evidence which may relate the performance of near visual tasks to progressive myopia. While this is not specific to VDUs, it may be aggravated by the long-term use of VDUs, especially under suboptimal conditions.

- There is no radiation hazard from VDUs.

- There is considerable need for further research on the human factors issues involved in VDU use and for the development of human factors standards and guidelines in such areas as display standards; illumination; workstation paramters; task design; work schedules; visual performance as related to VDU utilization; workload; task complexity.

The conclusions of the National Research Council were basically negative, in that they found no evidence indicating that there were any unusual risks to the use of VDUs in the workplace. These conclusions were qualified by noting the large void in our knowledge about the use of VDUs. A recent review article in the Human Factors Review summarizes the research in several areas related to the use of VDUs (Helander et al., 1984). this article summarized research in five areas: visual discomfort; anthropometry and biomechanics; work organization and job satisfaction; character and display design; lighting and reflectance. The authors found data on the independent variables shown in Table 1 and the dependent variables shown in Table 2.

The interested reader is referred to the original paper for a mass of detail on the methodologies used, the variables studied, and the results obtained. We will summarize the main points concerning each of the main research areas.

Visual Effects. Visual effects have been measured in two ways, by the measurement of visual functions, and by the collection of questionnaire and survey results. VDU users report more eye strain, eye fatigue, and general discomfort than non-VDU users. "Creative" users have significantly fewer problems than data-entry clerks. Since the data entry clerks rarely look at their VDU, it is likely that the frequency of complaints is due to the monotony and fixed postural requirements of their job.

Visual acuity is known to change with age, and there is limited evidence that the performance of visually demanding tasks may hasten this change. Temporary myopia develops as a workday progresses, but recovery is complete after 30 min. of rest. None of these effects is specific to the use of VDUs, but may occur in any demanding and close visual task.

Anthropometry and Biomechanics. The design of workstations, and furniture for VDU users is important and frequently neglected. Work posture and eye-to-task distances are much less variable for many VDU users than for non-VDU users. One possible remedy for this is the design of tasks or furniture that force the worker to move around or at least allow them to change their work posture and orientation. Table 3 shows the parameters recommended for VDU equipment and furniture. The general consensus of the literature reviewed by these authors is that furniture should be adjustable in most dimensions and that the VDU should be equipped with a detachable keyboard.

(Insert Fig 3. about here)

Work Organization and Job Satisfaction. The evidence suggests that various complaints of visual and physical discomfort from users of VDUs are primarily due to the design of their tasks and job. Job tasks should be designed to promote postural changes and adjustments, variations in eye-to-task distances, and variety in the work.

The biggest factor in job and task design for VDU users may be software. While a great deal of research is required in this area, software should be designed to organize, prompt, and pace the work in desirable ways.

Character and Display Design. Dark characters on a light background are preferable to more standard video displays. Small characters, and more densely packed text may be best for reading tasks in which meaningful text must be scanned and comprehended quickly. Larger characters are better for search and other tasks

where detection and recognition of individual characters and symbols are important.

Minimal spacing between characters horizontally is the width of one dot of the dot-matrix which makes up the screen display and the characters. Vertical spacing should be a minimum of 4 dots.

Much of the research summarized in this 1984 review article was conducted before some of the high-resolution graphic workstations became available, so that there is an open area of research on font design, character size, and other variables with these new types of workstations. Recommendations based on literature up through 1984 are that characters should be composed of square dots, with matrix a minimum of 5x7, with 7x9 or 9x11 preferred. In terms of visual angle, the minimum size of characters should be 16-18 min. In terms of screen size, the minimum, size of character should be 2.54 mm.

Lighting and Reflectance. Illumination level in the workplace should be 200-300 lux if the VDU user must use hardcopy source documents as well as the VDU. If use of source documents is not necessary, then a minimum of 100 lux is acceptable. A common standard which requires the screen-surround luminance contrast to be less than 1:3 is probably too strict, but would result in acceptable workstations. Specular reflections contribute to performance decrements, but screen treatments to eliminate them degrade the image. A slightly etched screen, which will reduce glare and reflection without noticeably degrading the image is a good compromise.

Results of Interviews with Image Analysts

During a field exercise, the author had an opportunity to interview 6 image analysts and two of their supervisors (the imagery librarian and a warrant officer in the image analysis field). These imagery analysts were stationed in a field-deployed image interpretation facility where they had access to both hardcopy film images and softcopy digital images and frequently had a choice between use of hardcopy or softcopy imagery exploitation.

The most striking result of interviews and observations of the work of these analysts was their strong preference for doing hardcopy image exploitation, even though they had two experimental softcopy image processing workstations available to them. Our interviews with these individuals were done to try to determine some of the factors which entered into this preference and to determine the implications of these factors for image processing workstation design.

Observation of the same analyst doing both hardcopy

exploitation and softcopy exploitation was consistent with the literature on VDU utilization summarized above. In hardcopy exploitation, the analyst used a light table, but also held the image up, moved it closer or further away from his or her eyes, and carried the image film from one place to another. Occasionally the analyst used magnifying equipment. Under special conditions he or she could use stereo viewing. In softcopy exploitation, the analyst sat erect facing the screen of the workstation, kept a relatively constant image-to-eye distance, and maintained position and posture within close limits. Certain aspects of softcopy exploitation were quite time consuming, involving substantial delays and waiting periods for the analyst while the computer performed certain tasks.

We developed several hypotheses to account for the very strong preference these analysts expressed for hardcopy exploitation. These hypotheses are listed and discussed in the following paragraphs, along with engineering implications.

Temporal Delays. One of the most common complaints of the analysts regarding the softcopy workstation they used concerned lengthy delays involved in very simple operations. The largest delays were involved in file operations, during which the analyst loaded an image into memory for exploitation. This delay was magnified if the analyst made an error anbd loaded the wrong file, or if the analyst was uncertain about which image to work with and had to load several files. Other noticeable delays were involved in various image manipulations, such as contrast enhancement, filtering, and sub-image formation.

Most of these delays could be easily remedied by upgrading the hardware and software used. The system used by these analysts did not make use of state-of-the-art peripheral storage systems, and appeared to make use of file management systems developed for text file handling. Specialized software and hardware for managing image files and for displaying images and their modifications would eliminate many of the delay problems. Finally the image file system interface could be improved to reduce the possibility of mistakes in image selection.

Observing Behavior. Observing behavior is defined by psychologists as behavior which has the consequence of obtaining or clarifying stimulus situations in the persons environment (Holland, 1958). Much of the work of the analyst is observing behavior. In hardcopy exploitation, the observing responses are familiar, and occur in everyday situations for all people. They include moving the image closer or further away from the eye, holding it up to the light or moving it around on the light table. More specialized observing behavior used by analysts in hardcopy exploitation, such as the use of magnification devices or stereo viewing are also familiar even to most children without specialized training in imagery analysis. Finally, almost all of

the observing behavior occuring during hardcopy analysis has immediate results, with no temporal delays.

Softcopy exploitation is quite different. In order to clarify something seen in the image on the screen, the analyst uses novel observing behavior, including software and hardware functions controlled from the workstation keyboard, trackball, or mouse. "Natural" observing responses are non-functional, since moving close to the image is usually not very productive, and since reorienting the image requires the use of software and hardware functions rather than handling the image.

The engineering implications of this hypothesis require further research. In the system actually observed, the main problem with the software- and hardware-based observing responses was the delays involved. The analyst had available software and hardware functions which were equivalent to magnifying the image, reorienting it, stretching or reducing the illumination levels and the contrast. But each of these operations took time to do and to undo if it proved not to be helpful. The time required to do these types of operations must be reduced in an operational workstation.

It is also the case that the hardware- and software-based observing behavior differs from "natural" or more familiar behavior used to clarify visual situations. It is possible that intensive training with softcopy analysis can make these responses more natural. It is also possible that long-term familiarity with computer systems will result in people who regard these types of responses as natural. If so, then recruiting and utilizing people with such experience would help.

Visual Effects. Analysts using softcopy analysis, like the office workers mentioned in the reviews surveyed above, maintain a relatively fixed posture and relatively fixed eye-to-image-display distance during exploitation. It is possible that this, combined with the visual characteristics of the display results in some type of visual discomfort which the analyst finds unpleasant and which results in reduced levels of performance. Helander et al. (1984) reported results which showed temporary loss of visual acuity and related problems with close visual work. It is possible that the conditions of softcopy analysis magnify these effects so that they are more noticeable than they are in hardcopy analysis.

The engineering implications can be understood only after further research. At the very least, workstations should be fully adjustable, and should encourage the analyst to vary eye-to-image-display distance. Multiple screens of different sizes and distances from the analyst might reduce visual problems, as would combinations of hardcopy and softcopy workstations.

Attention, Effort, and Workload. It is most likely that any visual problems associated with softcopy exploitation exert their effects on the analyst's preference by increasing the required amount of "mental effort" or the perceived workload of the analyst during processing. One of the visual effects which occurs with use of text processing VDUs is spatial frequency adaptation (Lunn & Banks, 1986). After prolonged viewing of material with high power at certain spatial frequencies, the viewer adapts to these frequencies (Levinson & Frome, 1979; Frome, Levinson, Danielson, Clavadetscher, 1979; Blakemore & Sutton, 1969), and has more difficulty in processing such material. (Spatial frequency is related to the perceived size of visual objects. The characters are all roughly the same size on a VDU so they present the user with a very regular stimulus. The visual system adapts to stimuli in this size range and requires the user to exert more "mental effort" and attention to his or her task. Stimuli of a different perceived size will be seen as usual.) According to Lunn and Banks, this spatial frequency adaptation would account for many of the visual effects observed, including the temporary myopia (Helander et al., 1984).

The engineering implications of this are similar to those for the visual effects. The visual display must allow variation in the spatial frequency bands presented to the analyst. It should also allow or even encourage movement by the analyst. The job should be designed so that analysts have frequent breaks from visual tasks.

All of these hypotheses require more extensive research. Many of them are based on observations of image analysts using softcopy workstations, but others are largely derived from research on office workstations, designed primarily for text processing and may or may not apply to image processing.

EXISTING IMAGE PROCESSING WORKSTATIONS

The literature reviewed above showed basically that there are no insurmountable problems in the use of softcopy imagery exploitation. But there are some problems which need to be addressed before softcopy exploitation will be both acceptable to the analyst and as effective as hardcopy exploitation.

We know that softcopy imagery analysis is currently performed by scientists and technicians working in such diverse areas as microscopic specimen analysis, medical imaging and radiology, and geophysical analysis of satellite images. In order to determine what a tactical imagery analysis workstation might look like, we reviewed the state of the art in imagery analysis workstations which were commercially available in each of the above fields.

We selected the names of every vendor who was listed in the AAAS Science "Guide to Scientific Instruments" under the heading "image processing" and requested information about their products. While many of the products were clearly inappropriate for application in military imagery analysis, it was possible to put together a composite of what a state of the art workstation would look like and to identify areas in which further research is required before such a workstation can be fielded. The names of the vendors whose products were reviewed are in the appendix, along with a summary of their products.

The vendors whose products were clearly appropriate for military tactical imagery analysis applications included the following:

- Comtal/3M
- MegaVision
- Microscience
- Quantex Corporation
- Symbolics Computer.

There were several other companies whose products are relevant and which were included in this review: Sky Computers, Inc.; Sun Microsystems, Inc.; and Tektronics. Based on a review of the products offered currently by these companies we have tentatively put together the following specifications for a prototype image analyst workstation.

System Overview. The system should be a compact workstation with two displays, one for the command interface and one for the high resolution color image display. The user should be able to bring an image onto his or her screen, and to perform any operation with a delay which takes no longer than 1/30 sec. The operator should have access to the system through a standard alpha-numeric keyboard, keypads and specific function keys, and some type of pointing device. For research and development purposes, the prototype workstations should have several pointing devices, including mouse, trackball, joystick, and possibly others. Furniture and system components should be adjustable within the parameters shown in Fig. 3 above.

System Hardware. Given the specification above, the hardware for image management and processing will have to include specialized processors. We believe that the image memory should be able to work with images of 1024x1024x8bit size, and at least 4 bands of spectral information. It would be preferable to be able to work with many images of this size. To manage such imagery, vector, array, or pixel data-flow processors will have to be used. This is within the state of the art. The specialized hardware should be capable of interfacing with a variety of host systems, including the PC/AT, Dec, and others so that flexibility is ensured in eventual system development.

Certain functions should be built into the image management hardware, including roam, zoom, and image operations like averaging, addition, subtraction, logical (boolean) functions, rotation, reflection, and warping and registration. Some of these operations may have to be done with specialized combinations of hardware and software.

Software System. Software will have to include a programming environment for research and development purposes, but we will not attempt to specify that here. As noted, some of the functions listed above may have to be constructed from specialized software rather than being "built in" to the hardware. Other functions which will probably have to be added in software are image file management software, display software management software, image processing software, and image measurement software.

Image file management. As we noted above, one of the important tasks that the image analyst has to perform in responding to an Exploitation Request is finding and exploiting the correct image in a library of imagery. In existing softcopy exploitation systems, this is a very time consuming operation. Imagery storage and imagery storage management will have to be constructed so that there are good directories of the "library", presented to the user in useful form, probably as menus, organized in some operationally useful way. The imagery library will also have to be very fast, so that loading an image takes no longer than 1/30 sec.

Display management. In addition to the display functions which we recommended be built into hardware, the user is probably going to want to select image regions for special processing. The user interface has to include the capability to select regions using a pointing device, to perform operations on those regions independently of the rest of the image. In addition to standard image operations, regions of interest should be capable of copying, editing, annotating, and separate storage for subsequent reference. It should be possible to warp and stretch regions of interest so they can be registered to other regions.

Image processing. At the present time, there is little information about the utility of most automated image processing functions in tactical image analysis. For research and development purposes, we believe that the following image processing functions should be available to the analyst:

- contrast manipulation capabilities to include increase and reduce, special characteristic functions such as logarithm and expontential. These should be mouse controllable through direct manipulation interface;

- digital filters to include high-pass, low-pass, median,

Laplacian, and user developed filters, with user selectable box-size to include 3x3, 5x5, and 7x7 at a minimum.

- special image processing including erosion, dilation, edge enhancement, special edge detectors such as Roberts and Sobel.

Image measurement. The analyst should have the capability to perform automatic measurement on features of the image. Software functionality should include centroid identification, moments, areas, intensity, optical density, length, angle, perimenter, surface area, and region histogram and statistical calculations.

Report preparation and annotation. Software should support text processing using forms generation to assist the analyst in preparation of RECCEXREPS and other reports, and to assist the analyst in annotation of imagery for future use.

Soldier-Machine Interface. All of the functionality of the system should be available to the analyst using pointing devices, menu structured commands, with minimal requirement to use the alpha-numeric keyboard (only during composition tasks such as annotation and for report preparation). Image display capabilities have already been described above. For the character display, a high-resolution monochrome display, with user interactions based on menus and windows is recommended. A number of standards for the design of interfaces using such displays are available for reference e.g. Schneiderman, 1986 or guidelines from Apple computer.

Finally, we believe that a system which approaches these specifications can be built primarily from off the shelf components which are available right now. It probably does not have to be custom built.

PLANS FOR FUTURE RESEARCH

The most important component of an image intelligence system for tactical deployment is the individual image analyst at his or her workstation. The value of the whole system depends on how well the analyst can do his or her job, on how accurately they can detect, identify, and recognize objects of importance in the imagery they receive. The value of the whole system also depends on how rapidly they can do their job. We believe that softcopy exploitation workstations offer the analyst the capability to process imagery both more accurately and more quickly than ever before. But the workstation and the work environment will have to be designed very carefully to allow the analyst to take full advantage of the hardware and software capabilities. And the other human aspects of the system, including personnel selection, training, and job and task assignment, will also have to be

carefully designed. We propose a large scale research project to attempt to quantitatively relate individual image analyst performance in terms of speed and accuracy to characteristics of the tasks, workstation, work environment, and personnel in the system. The research we are proposing involves the use of sequential research designs to identify the most important factors involved and then to refine a mathematical model of performance as it relates to these factors.

At the present time we are attempting to identify the variables involved. The literature reviews summarized in the first part of this paper provided some leads, but we will need the advice of experts in the field before we finally identify variables which we will include in our experiments. We have tentatively identified the following "clusters" of variables as determinants of analyst performance:

- personnel variables
- amount of training
- amount of experience
- MOS and Army history
- intelligence test scores
- educational background
- physical, and especially visual characteristics
- work environment variables
- furniture and workstation characteristics
- workstation location and environment
- workload
- shift characteristics
- time on the job (fatigue)
- stress
- time demands
- job task characteristics
- general task variables (e.g. measurement vs analysis)
- specific task variables (e.g. scanning vs specific target search)
- decisions or judgments required of the analyst
- purposes of the analysis
- source and type of imagery (e.g. radar vs video or electro-optical).
- imagery and characteristics
- noise
- image quality
- physical characteristics of image (e.g. spatial frequency characteristics, contrast)
- ecological characteristics of the image (what's in it? textures, surfaces, types of objects)
- chromaticity (monochrome imagery, multispectral imagery)
- image compression/restoration characteristics
- image manipulation capabilities
- software and hardware functionality and speed

- soldier machine interface
- image and character display characteristics
- pointing and user interaction devices types
- display management software functionality
- peripheral storage characteristics and interface

We believe that for many of these variables, the optimal values are already known and we need not include these in time consuming experimentation if we can identify them. The planned sequence of this research is as follows:

Review literature	This is partially done and can be complete by Jan 1988.
Consult with experts	This involves discussing the variables listed with experts at Ft. Huachuca and other locations. Time frame is between Jan '88 and April '88.
Identify existing data bearing on the issues	Exactly parallel above
Identify facilities and locations for experimentation	This will be done between Jan 88 and Jun 88.
Complete development of first part of sequential design (screening experiment)	This requires finalization of the variables and selection research of values for experimentation, and development of analysis plan. Done and documented by Sep 88.

Finally, it is possible that certain aspects of this research can be facilitated by the procurement of one or more workstations for experimentation. The cost of the hardware ranges from $10K upward. Such a workstation would have to be located near possible participants in the research, most likely Ft. Huachuca.

The outcome of the research is intended to be a quantitative model of the performance of analysts performing certain tasks. We intend that the model will be of a type which can legitimately be used to select optimal or near-optimal combinations of the variables listed above in a field deployable image analysts workstation.

BIBLIOGRAPHIC MATERIAL FOR IMINT

Bloomfield, J. R., Beckwith, W. E., Emerick, J., Marmurek, H. H., Tei, B. E., and Traub, B. H. (1978). Visual Search With Embedded Targets.Alexandria, VA.: U. S. Army Research Institute Technical Report TR-78-TH8.

Bonnet, D. G., and Snyder, H. L. (1978). Prediction of the Recognition of Real Objects as a Function of Photometric and Geometric Characeristics. Alexandria, VA.: U. S. Army Research Institute Technical Report TR-78-TH7.

Caelli, T. (1983). Energy processing and coding factors in texture discrimination and image processing. Perception and Psychophysics, 34(4), 349-55.

Caelli, T. (1982). On discriminating visual textures and images. Perception and Psychophysics, 31(2), 149-59.

Carswell, C. M., and Wickens, C. D. (1987). Information integration and the object display: an interaction of task demands and display superiority. Ergonomics, 30(3), 511-527.

Clarke, F. R., Welch, R. I., and Jeffrey, T. E. (1974). Development of a Psychophysical Photo Quality Measure. Alexandria, VA.: U. S. Army Research Institute NTIS AD-776 369.

De Corte, W. (1986). Optimal colors, phosphors, and illuminant characteristics for CRT displays: the algorithmic approach. Human Factors, 28(1), 39-47.

Ellis, S. R., and Stark, L. (1986). Statistical dependency in visual scanning. Human Factors, 28(4), 421-438.sFloru, R., Cail, F., and Elias, R. (1985). Psychophysiological changes during a VDU repetitive task. Ergonomics, 28(10), 1455-1468.

Frome, F. S., Levinson, J. Z., Danielson, J. T., and Clavadetscher, J. E. (1979). Shifts in Perception of Size after Adaptation to Gratings. Science, 14 December, 206, 1327-1329.

Gellatly, A. R. H. (1983). the function of spatial frequency analysis: test of a proposal. Perception and Psychophysics, 34(3), 301-304.

Gibson, J. J. (1979). The Ecological Approach to Visual Perception. Boston, MA.: Houghton Mifflin, Co.

Gibson, J. J. (1966). The senses Considered as Perceptual Systems. Boston, MA.: Houghton Mifflin Co.

Gould, J. D., Alfaro, L., Barnes, V., Finn, R., Grischkowsky, N., and Minuto, A. Reading is slower from CRT displays than from Paper: attempts to isolate a single-variable explanation. Human Factors, 39(3), 269-299.

Headquarters, Department of the Army (1985). Imagery Intelligence. Washington, DC: Department of the Army, Field Manual 34-55.

Headquarters, Department of the Army (1984). Soldier's Manual MOS 96D Imagery Analyst Skill Level 1. Washington, DC: Department of the Army, STP 34-96D1-SM.

Helander, M. G., Billingsley, P. A., and Schurick, J. M. (1984). An evaluation of Human Factors Research on Visual Display Terminals in the workplace. In F. A. Muckler (Ed.), Human Factors Review. Santa Monica, CA: The Human Factors Society, pp. 55-129.

Howard, J. H. Jr. (1986). Spatial Scale in Image Detection and Recognition. Arlington, VA.: ONR Technical Report ONR-86-25.

Hughes, H. C., Layton, W. M., Baird, J. C., and Lester, L. S. (1984). Global precedence in visual pattern recognition. Perception and Psychophysics, 35(4), 361-371.

Hutchins, E. L., Hollan, J. D., and Norman, D. A. (1985). Direct Manipulation Interfaces. La Jolla, CA.: UCSD Institute for Cognitive Science Report #8503.

IBM (1984). Human Factors of Workstations with Visual Displays. San Jose, CA.: IBM Human Factors Center.

Jeffrey, T. E., Martinek, H., Shvern, U., and Johnson, E. M. (1980). ARI Image Interpretation Research: 1970-1980. Alexandria, VA.: U. S. Army Research Institute Research Report 1252.

Julesz, B., & Papathomas, T. V. (1984). On spatial-frequency channels and attention. Perception and Psychophysics, 36(4), 398-399.aKroemer, K. H. E. (1983). Ergonomics of VDU Workplaces. Digital Design, Feb., 25,31-34.

Levinson, J. Z., and Frome, F. S. (1979). Perception of size of one object among many. Science, 21 December, 206, 1425-1426.oLo, C. P. (1986). Applied Remote Sensing. New York: Longman, Inc.

Lunn, R., and Banks, W. P. (1986). Visual Fatigue and Spatial Frequency Adaptation to Video Displays of Text. Human Factors, 1986, 28(4), 457-464.

Luria, S. M., Neri, D. F., and Jacobsen, A. R. (1986). The effects of set size on color matching using CRT displays. Human Factors, 28(1), 49-61.

Montgomery, C. A., Thompson, J. R., Katter, R. V. (1980). Imagery Intelligence (IMINT) Production Model. Alexandria, VA.: U. S. Army Research Institute Resarch Report 1210.

Matthews, M. L. (1986). The influence of visual workload history on visual performance. Human Factors, 28(6), 623-632.National

Research Council (1983). Video Displays, Work, and Vision. Washington, D. C.: National Academy Press.

Pearson, J. E., and Pearson, C. E. (1985). The effects of bandwidth compression and image quality on image interpreter performance. Human Factors, 1985, 27(3), 345-353.

Schneiderman, B. (1984). Beyond artificial intelligence: overcoming "the obstacle of animism." XXXX"

Schneiderman, B. (1983). Direct manipulation: a step beyond programming languages. IEEE Computer, 16(8), 57-69.

Schowengerdt, R. A. (1983). Techniques for Image Processing and Classification in Remote Sensing. New York: Academic Press.

Snyder, H. L. (1983). Visual ergonomics and VDT standards. Digital Design, Feb, 24, 26-30.

Wallman, J., Gottleib, M. D., Rajaram, V., Fugate-Wentzck, L. A. Local retinal regions control local eye growth and myopia. Science, 237, 3 July, 73-77.

Westheimer, G. (1984). Spatial vision. Annual Review of Psychology, 35, 201-26.3Williges, R. C. (1987). The use of models in human-computer interface design. Ergonomics, 30(3), 491-502.

APPENDIX: LISTING OF IMAGE PROCESSING EQUIPMENT VENDORS WHOSE PRODUCTS WERE REVIEWED DURING THIS RESEARCH

AMERICAN INNOVISION, INC.
7750 DAGGET ST.
SUITE 210
SAN DIEGO, CA 92111

This company offers two products of interest

- Videometric 150 for realtime color image acquisition and processing for medical application. The system has a small video camera which can capture a video image. It is linked to a PC/AT look-alike and a full-color video monitor. The system comes with software for image measurement and image processing, including automatic area measurement and counts of objects, brightness and color identification, and image statistics.

- Psychometry system for psychological experimentation. This system captures a video image which can be recorded on video-disk, and displayed at a PC/AT look-alike terminal equipped with a special touch-screen. Images can be displayed and a person's reaction to them recorded from the text screen. Software is provided for image capturing, editing, and for data recording during experimentation.

AMERSHAM CORPORATION
2636 SOUTH CLEARBROOK DRIVE
ARLINGTON HEIGHTS, IL 60005-4692

This company produces products for molecular biochemistry.

ARTEK SYSTEMS CORPORATION
SUBSIDIARY OF DYNATECH CORPORATION
170 FINN COURT
FARMINGDALE, NY 11735

This company provides image acquisition and analysis systems for the analysis of microscopic images in medical and metallurgical applications. The systems are based on a range of computers from Apple II and PC to minicomputer. Software is provided for a variety of imagery measurement and counting functions, as well as for imagery statistics.

BIOIMAGE
1460 EISENHOWER PLACE
ANN ARBOR, MI 48108A

This company makes two systems, both for laboratory analysis of two-dimensional electrophoresis patterns.

- The VISAGE system is based on a Motorola 68000 processor

with monitor and separate 1024x768 8-bit color monitor with hardware cursor control, roam and zoom. Software is provided for spot identification, area measurement, and true color analysis.a- The IQ system is a smaller version based on PC/AT.

BIOQUANT
R&M BIOMETRICS, INC.
5611 OHIO AVENUE
NASHVILLE, TN 37209N

This company provides a PC-based image acquisition and analysis system. Software provided includes analytic routines for both BW and Color image processing, measurement, and 3-D reconstruction and rotation.

COLORADO VIDEO, INC.
BOX 928
BOULDER, CO 80306B

Colorado Video provides a variety of hardware and software for the capture and analysis of "still video" photographs.

COLUMBUS INSTRUMENTS
P. O. BOX 44049
COLUMBUS, OH 43204C

This company provides very specialized hardware and software using video equipment to study the activity of laboratory animals.

COMTAL/3M
IMAGE PROCESSING SYSTEMS
1111 SOUTH ARROYO PARKWAY
PASADENA, CA 91105P

This company offers two basic image analysis systems, one based on PC, and the other based on mini-computer. This is probably as close as there is to the state of the art workstation for Army tactical image processing.

The PC based system offers a special image processor based on the Intel 80386 processor with both pseudo-color and true color capabilities. Hardware capabilities include a 512x512x8-bit image memory with real time roam, zoom, minify, and region of interest processing. A software package called image-pro provides image filtering, image statistics, image measurements, image operations such as addition, subtraction, and logical functions, and image editing.

The mini-computer based systems provides the same functionality with higher speed.

IMAGING RESEARCH INC.
BROCK UNIVERSITYST.
CATHARINES, ONTARIO, CANADA L2S 3A1S

This company markets an image capturing system and analysis system primarily for medical image analysis.

INDEC SYSTEMS, INC.
1283 MT. VIEW-ALVISO ROAD
SUNNYVALE, CA 94089S

This company provides image acquisition and analysis systems for medical image processing. Two systems, one based on Digital Equipment Corporation hardware, and one based on PC hardware are offered.

INFICON
LEYBOLD-HERAEUS INC.
6500 FLY ROAD
EAST SYRACUSE, NY 13057

This company markets specialized systems for medical fluoroscopic analysis.

MEGAVISION, INC
P.O. BOX 60158
SANTA BARBARA, CA 93160S

This company markets an image processing system which is close to the state of the art which is useful for military applications.

MICROSCIENCE
31101 18TH AVENUE SOUTH
FEDERAL WAY, WA 98003

This company offers real-time image acquisition and processing hardware and software. It is designed for applications including Satellite imagery and aerial photography.

MOTIONANALYSIS
93 STONY CIRCLE
SANTA ROSA, CA 95401S

This company markets CELLTRAK, a specialized image acquisition and analysis system for the study of cellular motility in medical and biological research applications.

NICOLET ANALYTICAL INSTRUMENTS
5225 VERONA ROAD
P.O. BOX 4508
MADISON, WI 53711-0508M

This company markets instruments for infrared spectroscopy.

ON-LINE INSTRUMENT SYSTEMS, INC.
ROUTE 2, BOX 111
JEFFERSON, GA 30549J

This company currently markets a digital image workstation specialized for images of the eye for use in medical research.

OPTOMAX
A DIVISION OF ITI, INC.
109 TERRACE HALL AVENUE
BURLINGTON, MA 01803B

This company markets an image processing and acquisition system based on PC for application in medical and laboratory image processing and in material science.

TED PELLA, INC.
4595 MOUNTAIN LAKES BLVD
REDDING, CA 96003

This company is a supplier of medical and biological research supplies, including image processing instruments.

PECEPTIVE SYSTEMS, INC.
1301 REGENTS PARK DRIVE
HOUSTON, TX 77058H

This company markets image acquisition and processing systems for medical applications with microscopic imagery.

PHOTONIC MICROSCOPY, INC.
2625 BUTTERFIELD RD,
#204-SOAK BROOK, IL 60531O

This company markets a general purpose image acquisition and processing primarily for microscopic image analysis.

PRINCETON GAMMA-TECH
A MEMBER OF THE OUTOKUMPU GROUP
1200 STATE ROAD
PRINCETON, NJ 80540

This company markets a general purpose image acquisition and processing system, primarily for analytical microscopy.

PROTEIN DATABASES INC.
405 OAKWOOD ROAD
HUNTINGTON STATION, NY 11744.

This company markets a workstation for two-dimensional gel electrophoresis.

QUANTEX CORPORATION
252 WOLFE ROAD
SUNNYVALE, CA 94086S

This company markets a general purpose image processing and analysis system based on PC/AT. This system has potential.

RAPID IMAGING SOFTWARE, INC.
P.O. BOX 160
TIJERAS, NM 87059

This company markets specialized software called the IMAGING TOOLKIT for image processing and construction for PC type computers.

SKY COMPUTERS
FOOT OF JOHN ST.
LOWELL, MA 01852

This company manufactures and markets special-purpose hardware for signal and image processing, including array processor boards.

GEORGE R. SNELL ASSOCIATES, INC.
269 SHEFFIELD STREET
MOUNTAINSIDE, NJ 07090

This company markets high-intensity projectors for use in viewing film images.

SOUTHERN MICRO INSTRUMENTS, INC.
120 INTERSTATE NORTH PARKWAY EAST
SUITE 308
ATLANTA, GA 30339A

This company markets instruments for image acquisition and analysis for use in microscopy.

SUN MICROSYSTEMS
2550 GARCIA AVENUE
MOUNTAINVIEW, CA 94043M

This company manufactures general purpose computer workstations, including some with high resolution color-graphics capabilities. These systems can be equipped with image processing software and possibly enhanced with high speed graphics hardware for use in image processing.

SURFACE SCIENCE LABORATORIES

1206 CHARLESTON ROAD
MOUNTAIN VIEW, CA 94043

This company markets specialized image acquisition and processing hardware and software for electron microscopy, infrared spectrometry, and x-ray image analysis.

There are no doubt many other vendors of image analysis products. We believe that there are a few in the above list who make products which represent the state of the art in image analysis workstations for tactical military image analysis and which provide us with a basis for specifying a prototype workstation.

Fig. 1. Independent Variables included in Helander et al. 1984 Review of Human Factors of VDUs.

Video Display Terminal Design

- Screen polarity and color
- Screen reflections and filters
- Screen position
- Screen luminance
- Character-Screen contrast
- Character design
- Text formatting
- Flicker
- Keyboard design

Office design

- Ambient illumination
- Screen-background contrast
- Furniture design
- Noise

Job Design

- Type of task
- Time on task

Subject Characteristics

- Age
- Body size
- Muscular disease

Fig 2. Dependent Variables included in Helander et al. (1984) review of Human Factors of VDU use.

Visual

- Self-reported strain
- Recognition-legibility
- Accommodation, vergence, acuity
- Eye movements
- Physiological measures

Musculoskeletal

- Self-reported strain
- Work posture
- Medical examination
- Physiological measures

Mental strain

- Self-reported strain
- Job satisfaction
- Physiological stress

Performance

- Time-error
- Measurement of workplace dimensions
- Furniture adjustment
- Preferred ambient, illumination
- Preferred character-screen contrast

Fig 3. Recommended Parameters for VDU Workstation (Helander et al., 1984).1

Screen surface to table edge		46-81 cm.
Screen viewing angle (below horiz.)		0-25 deg.
Keyboard angle		15 deg.
Footrest		14 deg.
Table height		
typing	70 cm.	
w/o keyboard	74 cm.	
Home Row height for keyboard		63-83 cm.
Chair height		38-50 cm.
Minimum spacing between characters		
horiz.	1 dot	
vert.	4 dot	
Minimum size of dot matrix for chars		5 x 7
Minimum size of char		16-18 min. visual angle
Minimum size of char		2.54 mm screen size
Illumination level	200-300 lux (with source document)	
Illumination level	100-200 lux (without source document)	

Where a range is indicated, equipment should be adjustable by the user.

Working Paper

WP HUA 89-02

Suggested Terra Scout Experiment Plan

George W. Lawton

January 1989

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

SUGGESTED TERRA SCOUT EXPERIMENT PLAN

CONTENTS

SUGGESTED TERRA SCOUT EXPERIMENT PLAN

Introduction

In the draft Terra Scout experiment plan which was submitted for our review (Appendix A), differences between the subjects and experimental conditions were completely confounded. There was no provision for identifying differences between the subjects that might exist prior to the experiment and to take account of these differences in evaluating differences between experimental conditions.

Suggested Design

A quasi-experimental design called a non-equivalent control groups design (Cook & Campbell, 1979) is more appropriate. This design can be implemented at little additional cost and will add immeasurably to the value of the experiment. In particular, it will allow identification of both differences between subjects and experimental conditions and assessment of these differences independently. The design is a quasi-experimental design, as opposed to a true experimental design due to the fact that there is no opportunity to use randomized assignment of treatments to experimental units to make the groups statistically equivalent.

In the adaptation of the non-equivalent control groups design which we suggest here, there are two stages. In the first stage, all subjects are asked to perform the same sequence of imagery tasks. This stage can be implemented during the preflight training at little additional cost. The power of the experiment would also be increased by adding subjects in the control group. Three would be a minimum. It is assumed that it is not possible to have more than a single subject in the experimental condition. While the addition of subjects is desirable, it is assumed in the rest of this paper that the experiment will be performed with two subjects, one in the experimental condition, and one in the control condition. Therefore, in stage 1, both subjects work a number of identical imagery problems using a ground facility which provides conditions, equipment, and imagery which are as equivalent as possible to those which will be used during the experimental condition. This sequence of problems will provide the experimenter the ability to assess differences between the

	Phase 1	Phase 2
Subject 1	MOE_{111}	$MOE_{12(n+1)}$
	MOE_{112}	$MOE_{12(n+2)}$
	MOE_{113}	...
	...	...
	...	...
	...	...
	MOE_{11n}	$MOE_{12(2n)}$
	$MOE_{11.}$	$MOE_{12.}$
Subject 2	MOE_{211}	$MOE_{22(n+1)}$
	MOE_{212}	$MOE_{22(n+2)}$
	...	...
	...	...
	...	...
	MOE_{21n}	$MOE_{22(2n)}$
	$MOE_{21.}$	$MOE_{22.}$

Table 1. Data layout for basic Terra Scout experiment. MOE_{ijk} is the measure of effectiveness for subject i, in condition j, and target or problem k, where i=1 or 2, j=1 or 2, and k=1 to n for phase 1 and (n+1) to 2n for phase 2, so that there are a total of 2n targets or problems in the experiment.

subjects prior to the actual experimental conditions, as well as providing a baseline against which the experimental data can be evaluated. Use of this design requires that the subjects be thoroughly practiced at the tasks and familiar with the equipment.

The imagery used in the first stage should have known values for Cloud Cover (CC) and Sun Angle or Time from Solar Noon (TFSN). The data collected during stage 1 should be as nearly equivalent as possible to the data which will be collected later during the experimental stage.

There are a number of factors which are confounded in the actual experiment but cannot be easily unconfounded. These factors include equipment (telescope vs. VDT), environment (ground vs. air platform), imagery (direct view vs. relayed digital), and probably others. These factors are completely confounded, so they have all been lumped into the condition variable called here "Phase". We thus have the following variables in the experiment: Subjects, CC, TFSN, Phase, Problem, and Measure of Effectiveness (MOE).

The data layout for this design is shown in Table 1. The non-equivalent control groups design may result in a number of possible outcomes. Figure 1 shows one hypothetical outcome. This figure shows the mean values of data in each of the cells in Table 1. The difference labeled D1 is a preexperimental difference between the experimental and control subjects. The difference labeled D2 is a difference between the experimental and control subjects during the actual experimental conditions, during which the experimental subject is aloft. The important question to be answered in the experiment is whether D1 is equal to D2. In analysis of variance terms, this is the interaction of groups or subjects with conditions, and the null hypothesis is that this difference is 0.

Suggested Analysis

Table 2 shows one way in which this hypothesis could be evaluated. It shows that the data could be reduced to sets of difference scores for each target, showing the difference between each subject's MOE for that target. Mean difference scores could be computed for each phase of the experiment and compared to evaluate the hypothesis of no difference. While conceptually very simple, this type of analysis has difficulty making use of the "covariates" data on Cloud Cover and Time from Solar Noon, which should increase the power of the experiment. Hence, I suggest that the data be analyzed using Multiple Regression and Correlation (MRC) analysis. This type of analysis will allow tests of the appropriate hypotheses while also making use of all the data, including covariates.

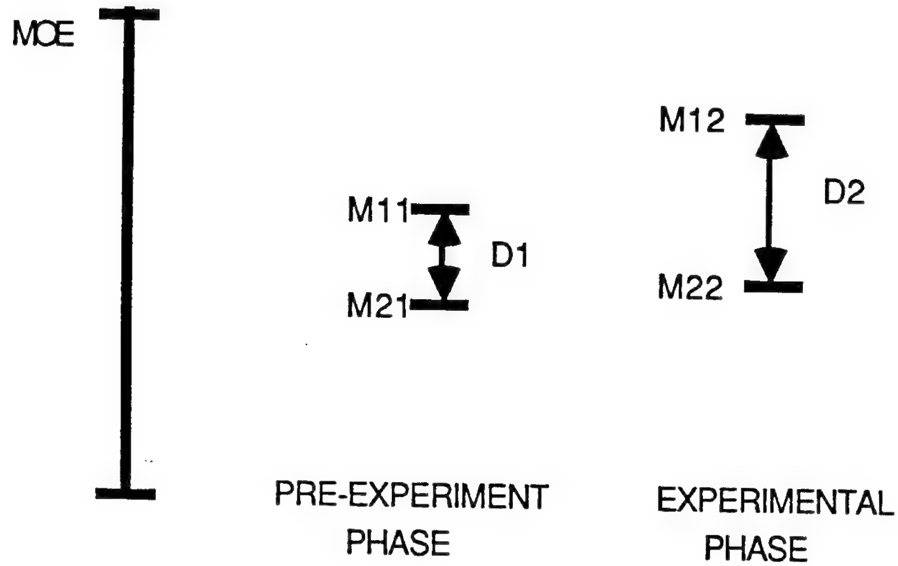


FIGURE 2. The figure shows a possible outcome of the non-equivalent control groups design with M11 representing the mean MOE for the experimental subject during phase 1, M12 representing the mean for the experimental subject during phase 2, and the M21 and M22 values representing the phase 1 and 2 values for the control subject.

Phase 1	Phase 2
$MOE_{211} - MOE_{111}$	$MOE_{22(n+1)} - MOE_{12(n+1)}$
$MOE_{212} - MOE_{112}$	$MOE_{22(n+2)} - MOE_{12(n+2)}$
...	...
...	...
...	...
$MOE_{21n} - MOE_{11n}$	$MOE_{22(2n)} - MOE_{12(2n)}$
D1	D2

Table 2. Possible analysis of Terra Scout Data. Since each subject works the same problems, an analysis could be performed on the differences between subjects for each phase of the experiment. Comparison would be between mean differences D1 and D2 for phases 1 and 2 (means of the columns of differences shown in the table).

Subj-Condition	Subject	Phase	SubjXPhase
Exp-Preflight	1	1	1
Con-Preflight	- 1	1	- 1
Exp-Flight	1	- 1	- 1
Con-Flight	- 1	- 1	1

Table 3. Contrast Coding for experimental conditions. Exp is experimental subject and Con is control subject.

Suggested Coding of Data for Experimental Conditions

The information in the experimental and subject variables should be coded using "contrast coding" (Cohen & Cohen, 1983). The observations made during the experiment fall naturally into four groups, as shown in Table 1. Three variables can be used to summarize these conditions. The values which should be used for each of these variables are shown in Table 3. The data layout which can be used with most statistical packages for regression analysis is exemplified in Appendix B. The layout is as follows: Problem, MOE, Subject, CC, TFSN, Phase, Subject X Phase interaction.

The variables in the experiment fall naturally into sets as follows: Set 1 - Subjects (between subject variability); Set 2 - Cloud Cover, Time from solar noon, Phase (within subject variability of no intrinsic interest); Set 3 - Subject X Phase interaction (within subject variability relevant to the hypothesis).

The hypothesis that the subject X Phase interaction is 0 can be tested in the MRC framework by computing the F statistic based on the ratio of variance proportions shown in equation (1). In (1), v is the degrees of freedom for the error term, which will depend on the number of problems in both conditions. The subscripts indicate the variable set number. Thus, each squared

$$F = ((R^2_{MOE.3,2} - R^2_{MOE.2})(1 - R^2_{MOE.3,2,1})^{-1}) * (v/1) \qquad (1)$$

multiple correlation term is the proportion of variance in MOE which is accounted by one or more of the variables in the numbered sets. The denominator of this ratio is the within-subject error variance.

The quantities which go into this F ratio can be obtained from a hierarchical or stepwise regression analysis in which the variable sets are entered in the following order: Set 2, Set 3, Set 1. Such an analysis will show the contribution of each variable or variable set over and above the variables which have already been entered into the analysis and will allow the computation of the appropriate differences and variance ratios. An example of such an analysis using the statistical package SPSS is provided in Appendix C.

Statistical Power of the Analysis

The power of the experiment will depend on the number of problems which are worked by the subjects in each phase of the experiment. I recommend that each subject work 20 problems in each phase of the experiment. This will give a total of 80 observations (20 for each of 2 conditions and for each of 2

subjects). This number of problems and the power were derived using the material in Cohen (1977, especially chapters 8 and 9 and Table 9.3.2 on p. 417) and Cohen & Cohen (1983, especially section 11.3.6, p. 449).

The rationale for chosing the number of problems given above depends on the following assumptions. For a MRC analysis of the type recommended here, Cohen (1977) expresses effect size (ES) in terms of a statistic called f^2. He identifies "small", "medium", and "large" effect sizes (ESs) as .02, .15, and .35 respectively. I have assumed that the Terra Scout experiment should be designed to detect "medium" size effects and thus used $f^2 = .15$. This is equivalent to $R^2 = .13$ or $R = .36$, where R is the multiple correlation value. Assuming an f^2 of .15, a = .05, u = 1, and total N = 80, and using the tables in Cohen (1977), one finds the power of the experiment to be .92 in the analysis of the subject X Phase interaction. Here a is the probability of type I error and u is the df for the Subject X Phase interaction.

Preliminary estimates of the number of problems required were obtained using the formula $N=/f^2+u+w+z+1$, where N is the total number of observations, L is a noncentrality parameter used by Cohen to table power, and u,w, and z are the numbers of variables in the three IV sets listed above, so that z=1 for the set 1, w=3 for the set 2, and u=1 for set 3. For u=1, a=.05, power =.90, L=10.51. Using $f^2=.15$, N=(10.51/.15)+6 which is 70. This implies 17.5 targets in each phase of the experiment to achieve power = .90. Requesting 20 targets will insure that the experiment has adequate power.

Interpretation of Experiment

While I have proposed the use of multiple correlation and regression analysis as the simplest and most elegant way to evaluate the Terra Scout data, and I have presented the power calculations based on the requirements of such an analysis, the experiment is concerned with mean differences. The effect size used, expressed as f^2, may not have the intuitive meaning that a standardized mean difference would to many experimenters. This ES can be algorithmically converted to a multiple correlation value R, which can in turn be used to approximate a t-statistic value for the test of the significance of the difference between two means using the formula $t=R/((1-R^2)/(n-2))^{-1/2}$.

Recall that an ES of $f^2=.15$ is equivalent to a multiple correlation value of R=.36. Substituting this value into the previous equation provides a t=2.38 with df=38. The df were arrived at by assuming that the appropriate t would be a t-test of differences between 2 mean differences based on 20 differences in each dondition. This value of t can be assumed to be a good approximation to a standarized normal deviate and should give the experimenter a feeling for the magnitude of mean difference

represented by the MRC ES.

REFERENCES

Cohen, J. (1977) Statistical Power Analysis for the Behavioral and Social Sciences. New York: Academic Press.

Cohen, J., and Cohen, P. (1983) Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cook, D., and Campbell, D. (1979) Quasi-Experimentation: Design and Analysis for Field Settings. Chicago: Rand McNally College Publishing Company.

APPENDIX A

ORIGINAL TERRA SCOUT EXPERIMENT PLAN

Draft Terra Scout Experiment Plan

Point of contact: CPT Apgar, autovon 879-8535

References: Army Space Operations Interim Operational Space Concept (U), TRADOC, 1985. (SECRET Document)

"Reports at Cosmonautics Lectures Honoring S.P. Korolev", *Leningradskaya Pravda*, 16 January 1982.

Purpose: This experiment's purpose is to evaluate the ability of an expert imagery analyst to conduct real time imagery analysis from a spacecraft in low earth orbit.

Subjects: The experiment will have two subjects. Each subject will be either a 96D30/40 noncommissioned officer or a 962A warrant officer. Both subjects will have at least ten years of imagery interpretation experience. This experience will include tours of duty in tactical and strategic intelligence. Each subject will have a minimum visual acuity of 20/20, or vision that is correctable to 20/20. Corrections must be made with contact lenses.

Subject training: Both subjects will receive extensive training on a simulation device. The exact number of training hours has not yet been determined. The simulation device will replicate the operation of a spaceborne telescope system which is focused on terrestrial targets. Specifically, the simulator will teach each subject to search, acquire, track, and observe targets which are

in range for approximately 70 seconds. The simulator will show each subject examples of targets which he will be expected to acquire during the experiment. The simulator cannot replicate a weightless environment.

Equipment: The spacecraft to be used in this experiment is the ---- ( ). The telescope to be used is the Spaceborne Direct-View Optical System (SPADVOS). The SPADVOS was designed and built by Dr. Lee Task of the Armstrong Aeromedical Research Laboratory. Some of its most important optical characteristics are as follows:

a) The objective lens will be a zoom lens from a 35mm camera.
b) Two zoom lenses are being considered; one has a 50 to 300mm focal length, and the other has a 120 to 600mm focal length.
c) The eyepieces will have a 20mm focal length.
d) The maximum telescopic magnification will be 40X.
e) Ground resolution is 15 feet.
f) The SPADVOS has a 45 degree viewing angle along the -----'s long-track and a 20 degree viewing angle along the -------'s cross-track.
g) When viewing straight down, the subject will be observing a ground swath with a 15 nautical mile diameter.

The SPADVOS' optical path includes a beamsplitter which allows a video camera to record the scenes viewed by the subject. Because of technical constraints, the video tape will only have a 50 foot resolution; however, it is still expected that this record will allow the experimenter to collect further data for analysis and modeling. The video tape will be exploited at the TENCAP Training Applications and Systems Integration Facility (TTASIF) at Ft. Huachuca.

Experiment plan: The concept of this experiment is to compare the quality of the imagery analysis done by the two subjects, and to use this comparison to determine if manned observation is of

value to the combined arms commander. Subject 1 will conduct imagery analysis from the ----. Subject 2 will conduct imagery analysis at the TTASIF. The imagery analysis done by each subject will be evaluated according to a specific measure of effectiveness (MOE). The results of these evaluations will be compared to make a judgement on the value of manned observation.

A set of targets will be generated by correlating areas of military interest with ---- ephemeris data. Once the targets are selected, they will be placed into a target notebook. Each subject will have the same target notebook. This notebook will consist of target imagery and maps of the surrounding areas. Accompanying each photograph will be a target description to assist the subject in acquisition. Finally, each target will have an associated task list. For example, if the target were an airfield, a task might be to count the number and type of jet aircraft present. Example target descriptions and task lists are contained in enclosures 1 and 2.

Procedure: After the ____ reaches orbit, the mission commander will tell Subject 1 when to begin the installation and checkout of the SPADVOS in accordance with the Crew Activity Plan (CAP). At ten minutes prior to the target, the -------- -------- -------- -------- (SPOC) will cue Subject 1 to conduct final preparations to observe the target. Subject 1 will then do a final equipment check. He will open the target notebook,look at the target imagery, read the target description, and review the tasks which he is to accomplish. At two minutes prior to the

target, the SPOC will cue Subject 1 to begin the search. Subject 1 will then search, acquire, track, observe, and accomplish the listed tasks. He will continuously describe his actions so as to create a record of his thoughts on the video tape's audio channel. Fifteen minutes after the second cue, Subject 1 will cease work, and make final enties in the task section of the target notebook. This procedure will be repeated for all targets in the notebook.

After the ---- lands, the video tape will be recovered and sent to the TTASIF where the tape will be exploited using computer enhancement techniques. The TTASIF will also receive imagery of the same targets observed from the -------.This imagery will be collected by conventional methods at as close to the time as the targets were observed from the ------- as possible. Subject 2 will have fifteen minutes in which to use his target notebook to exploit this conventionally-collected imagery.

Results: An independent committee of expert imagery analysts will evaluate each subject's task performance according to one MOE. The committee will not know which subject flew on the ----. The MOE will be Response Accuracy. This MOE is defined as:

Response Accuracy - Each task response will be graded on accuracy and completeness on a percentage basis. This number will represent the percent correct of each task response. If a target is not acquired because of cloud cover or any other reason, the response accuracy will be 0%.

In addition to this performance standard, the independent

committee will record cloud cover and the time from solar noon for each target. These parameters will be recorded to verify the effects of cloud cover and sun angle on imagery analysis. They are defined below:

a) Cloud Cover - The committee will study the video tape and make a subjective estimate of the cloud cover in the target vicinity to the nearest 5%. Percent cloud cover will be estimated for the atmosphere two minutes prior to the target to two minutes after the target.

b) Time From Solar Noon - The committee will record the time from solar noon that each target was attempted. Solar noon at the median target longitude will be used for all calculations.

After completing the evaluation, the independent committee will record the results according to the general format contained in enclosure 3.

Planned analysis: The experimenter at USAICS Space Division will analyze the data collected according to enclosure 3. He will determine the statistical significance of the difference in the imagery analysis conducted by the two subjects. The experimenter must use the MOE to determine whether manned observation's speed is more valuble to the combined arms commander than the expected higher quality of ground-processed imagery. It is projected that this determination will be made with the assistance of mathematical modeling.

Target Description

Target 1 is a military airfield. A rural road, which leads to the airfield, intersects a major highway 5 miles west of the airfield. A small town of an approximate 3 mile radius is located at this intersection.The airfield itself has two runways oriented from east to west, and one runway oriented from north to south. Three hangers and the control tower are positioned to the west of the north/south runway.

Target Task List

1. Identify and count the number of jet aircraft parked at the airfield.

2. Identify and count the number of helicopters parked at the airfield.

3. Determine which of the hanger doors are open, and which are closed.

Results Record

Subject 1

Target #	Cloud Cover	Sun Angle	Response Accuracy %

Subject 2

Target #	Cloud Cover	Sun Angle	Response Accuracy %

enclosure 3

APPENDIX B

DATA LAYOUT AND SIMULATED DATA

```
1 48 -1 43 1708 1 -1
2 50 -1 40 0349 1 -1
3 48 -1 41 0734 1 -1
4 52 -1 44 1231 1 -1
5 50 -1 36 0706 1 -1
6 47 -1 38 1350 1 -1
7 52 -1 40 1450 1 -1
8 44 -1 41 1133 1 -1
9 53 -1 45 1548 1 -1
10 48 -1 45 1924 1 -1
11 52 -1 41 2153 1 -1
12 39 -1 34 1745 1 -1
13 49 -1 46 0550 1 -1
14 49 -1 34 0440 1 -1
15 50 -1 41 1853 1 -1
16 55 -1 45 0933 1 -1
17 44 -1 44 1320 1 -1
18 41 -1 43 0034 1 -1
19 49 -1 46 1303 1 -1
20 51 -1 40 1752 1 -1
21 47 -1 52 0808 -1 1
22 50 -1 48 1412 -1 1
23 48 -1 48 2010 -1 1
24 48 -1 51 1658 -1 1
25 49 -1 51 2128 -1 1
26 56 -1 54 0729 -1 1
27 54 -1 47 2305 -1 1
28 47 -1 53 2140 -1 1
29 50 -1 50 1627 -1 1
30 50 -1 51 1428 -1 1
31 45 -1 49 1404 -1 1
32 56 -1 54 2347 -1 1
33 52 -1 55 1653 -1 1
34 42 -1 57 0336 -1 1
35 57 -1 52 2128 -1 1
36 47 -1 51 2216 -1 1
37 54 -1 52 2141 -1 1
38 45 -1 55 0512 -1 1
39 50 -1 55 1653 -1 1
40 53 -1 52 0321 -1 1
1 54 1 43 1708 1 1
2 44 1 40 0349 1 1
3 48 1 41 0734 1 1
4 49 1 44 1231 1 1
5 50 1 36 0706 1 1
6 51 1 38 1350 1 1
7 49 1 40 1450 1 1
8 58 1 41 1133 1 1
9 53 1 45 1548 1 1
10 51 1 45 1924 1 1
11 58 1 41 2153 1 1
12 48 1 34 1745 1 1
13 52 1 46 0550 1 1
14 48 1 34 0440 1 1
15 42 1 41 1853 1 1
16 48 1 45 0933 1 1
17 48 1 44 1320 1 1
18 49 1 43 0034 1 1
19 51 1 46 1303 1 1
20 47 1 40 1752 1 1
21 59 1 52 0808 -1 -1
22 60 1 48 1412 -1 -1
23 54 1 48 2010 -1 -1
24 53 1 51 1658 -1 -1
25 52 1 51 2128 -1 -1
26 58 1 54 0729 -1 -1
```

This is the data file used to simulate Terra Scout results. It was produced with a random number generator and the coding scheme ("contrast coding") described in the paper.

Data layout is, by column
1 Prob 2 MOE 3 Subj 4 CC 5 TFSN 6 Phase 7 SXPhase

27	49	1	47	2305	-1	-1
28	67	1	53	2140	-1	-1
29	63	1	50	1627	-1	-1
30	53	1	51	1428	-1	-1
31	58	1	49	1404	-1	-1
32	58	1	54	2347	-1	-1
33	59	1	55	1653	-1	-1
34	60	1	57	0336	-1	-1
35	67	1	52	2128	-1	-1
36	60	1	51	2216	-1	-1
37	54	1	52	2141	-1	-1
38	62	1	55	0512	-1	-1
39	62	1	55	1653	-1	-1
40	61	1	52	0321	-1	-1

APPENDIX C

ANNOTATED EXAMPLE SPSS ANALYSIS OF SIMULATED DATA

```
18-Oct-88   SPSS-X RELEASE 3.0 FOR VAX/VMS
08:33:11    SPSS-X VAX/VMS Site            on RED::                    V4.7


VAX-11/780                 SPSS-X VAX/VMS Site            License Number 18569
This software is functional through September 30, 1988.

********************************************************************************

            Try the new SPSS-X Release 3.0 features:

* Interactive SPSS-X command execution                    * Improvements in:
* Online, VMS-like Help                                   *   REPORT
* Nonlinear Regression                                    *   TABLES
* Time Series and Forecasting (Trends)                    *   Simplified Syntax
* Macro Facility                                          *   Matrix I/O

See SPSS-X User's Guide, Third Edition, for more information on these features.

********************************************************************************

   1  0  FILE HANDLE     TERRA/NAME="TS.DAT"
   2  0  DATA LIST       FILE=TERRA LIST/PROB MOE SUBJ CC TFSN X1 X2
   3  0  VAR LABELS      MOE '% of targets identified'
   4  0                  SUBJ 'Analyst'
   5  0                  CC '% of cloud cover'
   6  0                  TFSN 'time from solar noon 24hr clock'
   7  0                  X1 'phase of experiment'
   8  0                  X2 'interaction contrast'
   9  0  VALUE LABELS    SUBJ -1 'Control - Ground-based'
  10  0                          1 'Experimental - Airborne'/
  11  0                  X1 1 'Pretest' -1 'Experimental Phase'
  12  0  PRINT FORMATS   MOE CC (F2.1) TFSN (TIME)
  13  0  MEANS           VARIABLES=SUBJ(-1,1) X1(-1,1) X2(-1,1) MOE(LO,HI)/
  14  0                  TABLES=MOE BY SUBJ/MOE BY X1/MOE BY X1 BY SUBJ

INTEGER BREAKDOWN NEEDS        504 BYTES OF MEMORY.

THERE ARE  4471424 BYTES OF MEMORY AVAILABLE.
```

18-Oct-88 SPSS-X RELEASE 3.0 FOR VAX/VMS
08:33:15 SPSS-X VAX/VMS Site on RED:: V4.7

D E S C R I P T I O N O F S U B P O P U L A T I O N S

Criterion Variable MOE % of targets identified
Broken Down by SUBJ Analyst

Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			51.7250	5.7147	80
SUBJ	-1	Control - Ground-bas	49.2750	4.0509	40
SUBJ	1	Experimental - Airbo	54.1750	6.1180	40

Total Cases = 80

This is a Table of means by subject

18-Oct-88 SPSS-X RELEASE 3.0 FOR VAX/VMS
08:33:15 SPSS-X VAX/VMS Site on RED:: V4.7

D E S C R I P T I O N   O F   S U B P O P U L A T I O N S

Criterion Variable MOE % of targets identified
Broken Down by X1 phase of experiment

Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			51.7250	5.7147	80
X1	-1	Experimental Phase	54.2250	6.1331	40
X1	1	Pretest	49.2250	3.9645	40

Total Cases = 80

Table of means for each phase of the experiment

18-Oct-88 SPSS-X RELEASE 3.0 FOR VAX/VMS
08:33:15 SPSS-X VAX/VMS Site on RED:: V4.7

D E S C R I P T I O N   O F   S U B P O P U L A T I O N S

Criterion Variable	MOE	% of targets identified
Broken Down by	X1	phase of experiment
by	SUBJ	Analyst

Variable	Value	Label	Mean	Std Dev	Cases
For Entire Population			51.7250	5.7147	80
X1	-1	Experimental Phase	54.2250	6.1331	40
SUBJ	-1	Control - Ground-bas	50.0000	4.0782	20
SUBJ	1	Experimental - Airbo	58.4500	4.7956	20
X1	1	Pretest	49.2250	3.9645	40
SUBJ	-1	Control - Ground-bas	48.5500	3.9931	20
SUBJ	1	Experimental - Airbo	49.9000	3.9189	20

Total Cases = 80

Tables showing means used to analyze the phaseXSubject interaction. We want to test the hypothesis that the interaction contrast is 0.

(-1)58.45 + (+1)50.0 + (+1)49.9 + (-1)48.55 =?? 0

	Pretest (+1)	Experiment (-1)	Means
Airborne Experimental (+1)	49.9	58.45	54.18
Ground Control (-1)	48.55	50.0	49.28
Means	49.23	54.23	

```
PRECEDING TASK REQUIRED       1.05 SECONDS CPU TIME;       1.72 SECONDS ELAPSED.

  15  0   REGRESSION      VARS=MOE TO X2/STATISTICS=ALL
  16  0                   /DEP=MOE/ENTER CC/ENTER TFSN
  17  0                   /ENTER X1/ENTER X2/ENTER SUBJ

THERE ARE  4472240 BYTES OF MEMORY AVAILABLE.

     2068 bytes of memory required for REGRESSION procedure.
        0 more bytes may be needed for Residuals plots.
```

18-Oct-88   SPSS-X RELEASE 3.0 FOR VAX/VMS
08:33:17   SPSS-X VAX/VMS Site   on RED::   V4.7

* * * * M U L T I P L E   R E G R E S S I

Listwise Deletion of Missing Data

Equation Number 1   Dependent Variable..   MOE   % of targets identified

Beginning Block Number 1. Method: Enter   CC

Variable(s) Entered on Step Number 1..   CC   % of cloud cover

Multiple R	.45791			Analysis
R Square	.20968	R Square Change	.20968	
Adjusted R Square	.19955	F Change	20.69391	Regressi
Standard Error	5.11282	Signif F Change	.0000	Residual
				F =

Condition number bounds: 1.000,   1.000

Var-Covar Matrix of Regression Coefficients (B)
Below Diagonal: Covariance   Above: Correlation

	CC
CC	.00875

XTX Matrix

	CC	MOE	SUBJ	TFSN	X1	X2
CC	1.00000	-.45791	.00000	-.15675	.85916	.00000
MOE	.45791	.79032	.43143	.12174	-.04682	-.31256
SUBJ	.00000	.43143	1.00000	.00000	.00000	.00000
TFSN	.15675	.12174	.00000	.97543	-.12920	.00000
X1	-.85916	-.04682	.00000	-.12920	.26185	.00000
X2	.00000	-.31256	.00000	.00000	.00000	1.00000

Cloud cover accounts for 20.97 percent of variance in MOE

```
                                       * * * *   M U L T I P L E   R E G R E S S I

Equation Number 1    Dependent Variable..   MOE    % of targets identified

----------------------------------------------------- Variables in the Equation -

Variable                 B          SE B      95% Confdnce Intrvl B        Beta     SE B

CC                 .425549       .093547        .239312       .611786    .457906     .100
(Constant)       31.894416      4.396594      23.141470     40.647362

------ in -------            ------------------ Variables not in the Equation ----

Variable     Sig T           Variable      Beta In   Partial   Tolerance   Min Toler

CC           .0000           SUBJ          .431425   .485292    1.000000    1.000000
(Constant)   .0000           TFSN          .124806   .138654     .975430     .975430
                             X1           -.178793  -.102914     .261851     .261851
                             X2           -.312563  -.351589    1.000000    1.000000

End Block Number   1   All requested variables entered.


                                       * * * * * * * * * * * * * * * * * * * * * * *

Beginning Block Number  2.  Method:  Enter      TFSN

Variable(s) Entered on Step Number  2..    TFSN      time from solar noon 24hr c

Multiple R              .47421                                              Analysis
R Square                .22487         R Square Change       .01519
Adjusted R Square       .20474         F Change             1.50934         Regressi
Standard Error         5.09621         Signif F Change        .2230         Residual

                                                                            F =

Condition number bounds:      1.025,      4.101
```

TFSN accounts for 1.5 percent ofvariance in MOE.

* * * * M U L T I P L E R E G R E S S I

Equation Number 1 Dependent Variable.. MOE % of targets identified

Var-Covar Matrix of Regression Coefficients (B)
Below Diagonal: Covariance Above: Correlation

	CC	TFSN
CC	.00891	-.15675
TFSN	-1.337E-05	8.162E-07

XTX Matrix

	CC	TFSN	MOE	SUBJ	X1	X2
CC	1.02519	-.16070	-.43834	.00000	.83839	.00000
TFSN	-.16070	1.02519	-.12481	.00000	.13245	.00000
MOE	.43834	.12481	.77513	.43143	-.03069	-.31256
SUBJ	.00000	.00000	.43143	1.00000	.00000	.00000
X1	-.83839	-.13245	-.03069	.00000	.24474	.00000
X2	.00000	.00000	-.31256	.00000	.00000	1.00000

-- Variables in the Equation -

Variable	B	SE B	95% Confdnce Intrvl B		Beta	SE B
CC	.407368	.094410	.219374	.595362	.438343	.101
TFSN	.001110	9.0343E-04	-6.89048E-04	.002909	.124806	.101
(Constant)	31.210745	4.417500	22.414378	40.007112		

------ in -------

Variable	Sig T
CC	.0000
TFSN	.2230
(Constant)	.0000

----------------- Variables not in the Equation ----

Variable	Beta In	Partial	Tolerance	Min Toler
SUBJ	.431425	.490025	1.000000	.975430
X1	-.125409	-.070468	.244738	.244738
X2	-.312563	-.355018	1.000000	.975430

End Block Number 2 All requested variables entered.

```
18-Oct-88    SPSS-X RELEASE 3.0 FOR VAX/VMS
08:33:18     SPSS-X VAX/VMS Site              on RED::                V4.7

                                       * * * *   M U L T I P L E   R E G R E S S I

Equation Number 1    Dependent Variable..   MOE   % of targets identified

Beginning Block Number  3.  Method:  Enter      X1


Variable(s) Entered on Step Number  3..    X1        phase of experiment


Multiple R             .47825                                                Analysis
R Square               .22872        R Square Change      .00385
Adjusted R Square      .19828        F Change             .37928             Regressi
Standard Error        5.11687        Signif F Change      .5398              Residual

                                                                             F =

Condition number bounds:      4.086,      27.240


Var-Covar Matrix of Regression Coefficients (B)
Below Diagonal:  Covariance    Above:  Correlation

                   CC         TFSN          X1

CC              .03416       .14174      .85846
TFSN         2.458E-05    8.803E-07      .25564
X1              .18348    2.774E-04     1.33727


XTX Matrix

                   CC         TFSN          X1  |        MOE  |        SUBJ          X2
                                                |             |
CC             3.89726       .29305     3.42568 |    -.33320  |      .00000      .00000
TFSN            .29305      1.09687      .54120 |    -.10820  |      .00000      .00000
X1             3.42568       .54120     4.08601 |     .12541  |      .00000      .00000
------------------------------------------------+-------------+------------------------
MOE             .33320       .10820     -.12541 |     .77128  |      .43143     -.31256
------------------------------------------------+-------------+------------------------
SUBJ            .00000       .00000      .00000 |     .43143  |     1.00000      .00000
X2              .00000       .00000      .00000 |    -.31256  |      .00000     1.00000
```

Phase (here called X1) accounts for less than 1 percent of
variance in MOE.

* * * * M U L T I P L E R E G R E S S I

Equation Number 1 Dependent Variable.. MOE % of targets identified

-- Variables in the Equation -

Variable	B	SE B	95% Confdnce Intrvl B		Beta	SE B
CC	.309656	.184821	-.058448	.677759	.333200	.198
TFSN	9.62186E-04	9.3827E-04	-9.06538E-04	.002831	.108195	.105
X1	-.712180	1.156403	-3.015356	1.590996	-.125409	.203
(Constant)	35.967908	8.907300	18.227482	53.708333		

------ in -------

Variable	Sig T
CC	.0980
TFSN	.3084
X1	.5398
(Constant)	.0001

----------------- Variables not in the Equation ----

Variable	Beta In	Partial	Tolerance	Min Toler
SUBJ	.431425	.491247	1.000000	.244738
X2	-.312563	-.355903	1.000000	.244738

End Block Number 3 All requested variables entered.

* *

Beginning Block Number 4. Method: Enter X2

Variable(s) Entered on Step Number 4.. X2 interaction contrast

Multiple R	.57133			Analysis
R Square	.32642	R Square Change	.09770	
Adjusted R Square	.29049	F Change	10.87790	Regressi
Standard Error	4.81361	Signif F Change	.0015	Residual
				F =

Condition number bounds: 4.086, 40.321

The subj X phase interaction accounts for 9.77 percent of variance in MOE. In MOE units this is (sign is not important)

(-1)58.45 + (+1)50.0 + (+1)49.9 + (-1)48.55 = -7.1

This is the statistic of interest. It is the difference of the differences between subjects in each phase of the experiment.

**** MULTIPLE REGRESSI

Equation Number 1 Dependent Variable.. MOE % of targets identified

Var-Covar Matrix of Regression Coefficients (B)
Below Diagonal: Covariance Above: Correlation

	CC	TFSN	X1	X2
CC	.03023	.14174	.85846	.00000
TFSN	2.175E-05	7.791E-07	.25564	.00000
X1	.16237	2.455E-04	1.18345	.00000
X2	.00000	.00000	.00000	.28964

XTX Matrix

	CC	TFSN	X1	X2	MOE	SUBJ
CC	3.89726	.29305	3.42568	.00000	-.33320	.00000
TFSN	.29305	1.09687	.54120	.00000	-.10820	.00000
X1	3.42568	.54120	4.08601	.00000	.12541	.00000
X2	.00000	.00000	.00000	1.00000	.31256	.00000
MOE	.33320	.10820	-.12541	-.31256	.67358	.43143
SUBJ	.00000	.00000	.00000	.00000	.43143	1.00000

-- Variables in the Equation -

Variable	B	SE B	95% Confdnce Intrvl B		Beta	SE B
CC	.309656	.173868	-.036706	.656017	.333200	.187
TFSN	9.62186E-04	8.8266E-04	-7.96163E-04	.002721	.108195	.099
X1	-.712180	1.087866	-2.879319	1.454959	-.125409	.191
X2	-1.775000	.538178	-2.847105	-.702895	-.312563	.094
(Constant)	35.967908	8.379385	19.275317	52.660499		

------ in -------

Variable	Sig T
CC	.0790
TFSN	.2792
X1	.5147
X2	.0015
(Constant)	.0001

------------------ Variables not in the Equation ----

Variable	Beta In	Partial	Tolerance	Min Toler
SUBJ	.431425	.525666	1.000000	.244738

* * * * M U L T I P L E R E G R E S S I

Equation Number 1 Dependent Variable.. MOE % of targets identified

End Block Number 4 All requested variables entered.

* *

Beginning Block Number 5. Method: Enter SUBJ

Variable(s) Entered on Step Number 5.. SUBJ Analyst

Multiple R	.71592			Analysis
R Square	.51254	R Square Change	.18613	
Adjusted R Square	.47961	F Change	28.25577	Regressi
Standard Error	4.12247	Signif F Change	.0000	Residual
				F =

Condition number bounds: 4.086, 55.401

Var-Covar Matrix of Regression Coefficients (B)
Below Diagonal: Covariance Above: Correlation

	CC	TFSN	X1	X2	SUBJ
CC	.02217	.14174	.85846	.00000	.00000
TFSN	1.595E-05	5.714E-07	.25564	.00000	.00000
X1	.11909	1.800E-04	.86801	.00000	.00000
X2	.00000	.00000	.00000	.21243	.00000
SUBJ	.00000	.00000	.00000	.00000	.21243

Subject accounts for 18.6 percent of the variance in MOE.

**** M U L T I P L E R E G R E S S I

Equation Number 1 Dependent Variable.. MOE % of targets identified

XTX Matrix

	CC	TFSN	X1	X2	SUBJ	MOE
CC	3.89726	.29305	3.42568	.00000	.00000	-.33320
TFSN	.29305	1.09687	.54120	.00000	.00000	-.10820
X1	3.42568	.54120	4.08601	.00000	.00000	.12541
X2	.00000	.00000	.00000	1.00000	.00000	.31256
SUBJ	.00000	.00000	.00000	.00000	1.00000	-.43143
MOE	.33320	.10820	-.12541	-.31256	.43143	.48746

--- Variables in the Equation -

Variable	B	SE B	95% Confdnce Intrvl B		Beta	SE B
CC	.309656	.148904	.012959	.606353	.333200	.160
TFSN	9.62186E-04	7.5593E-04	-5.44032E-04	.002468	.108195	.085
X1	-.712180	.931670	-2.568572	1.144213	-.125409	.164
X2	-1.775000	.460906	-2.693376	-.856624	-.312563	.081
SUBJ	2.450000	.460906	1.531624	3.368376	.431425	.081
(Constant)	35.967908	7.176272	21.668874	50.266942		

------ in -------

Variable	Sig T
CC	.0410
TFSN	.2071
X1	.4471
X2	.0002
SUBJ	.0000
(Constant)	.0000

End Block Number 5 All requested variables entered.

Systems Research Laboratory - Ft. Huachuca FU (Alexandria)

Working Paper #87-13

A COGNITIVE SCIENCE APPROACH TO ELICITATION OF EXPERT KNOWLEDGE

John M. Leddo

Marvin S. Cohen

Decision Sciences Consortium

Produced Under Contract #MDA903-86-C-0383

George W. Lawton

Contracting Officer's Technical Representative

August 1987

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

A COGNITIVE SCIENCE APPROACH TO ELICITATION OF EXPERT KNOWLEDGE*

John M. Leddo
Marvin S. Cohen

Decision Science Consortium, Inc.
7700 Leesburg Pike, Suite 421
Falls Church, Virginia 22043

ABSTRACT

This paper presents a framework for eliciting and representing expert knowledge based on theories of knowledge representation found in the cognitive science literature. Representation schemes for expert knowledge are constructed for both interpretative ("top-down") and generative ("bottom-up") approaches to structuring knowledge. Elicitation techniques for each approach are selected based on considerations of the knowledge needed to fill each representation scheme and the knowledge the techniques are hypothesized to be best suited for. These two approaches are tested between-subjects on twenty officers and enlisted men on active duty in the United States Army. The strengths and weaknesses of each approach and conditions under which each approach is likely to be most effective are discussed. Finally, applications of our framework to the problems of identification of experts, training, and the design of expert systems are discussed.

INTRODUCTION

The development of successful expert systems and Artificial Intelligence (AI) models of problem solving depends heavily on how well expert knowledge can be captured and represented in ways that are compatible both with how the expert him/herself represents knowledge and with the architecture within which the knowledge is to be embedded. It is our opinion that, traditionally, the process of knowledge elicitation has been guided more by considerations of expert system architecture than by considerations of how the experts actually represent knowledge. We feel that this both shortchanges the expert system that eventually gets constructed (since it is then a poor model of the expert) and also leads to an inferior knowledge elicitation process, which may involve incomplete or distorted output on the part of the expert.

Our approach is based on the argument that the process of eliciting and representing knowledge should be guided by considerations of how experts actually represent and use knowledge. These considerations guide both the selection of elicitation procedures and evaluation of the output produced using those procedures. We draw heavily from both cognitive psychology and AI theories of how people in general and experts, in particular, represent knowledge. We are particularly interested in expert problem solving, including the evaluation of situations, events, and incoming information in relation to specific goals,the generation of plans to achieve those goals, and the simulation of possible future events, especially outcomes relevant to the expert's goals.

A recurring theme in both the cognitive psychology and AI literatures is whether knowledge representation and information processing proceeds in a top-down or bottom-up fashion. By top-down (or interpretative as it is also called), we mean that knowledge is organized around high-level, expectancy-driven structures which form a preset framework for information processing. This preset framework dictates how incoming information is to be interpreted and what procedures are to be used in problem solving. By bottom-up (or generative), we mean that knowledge is organized around low-level rules and/or concepts, in the absence of a particular context. Hence, information processing and problem solving revolves around individual pieces of data or problem solving procedures and generates (hence the term generative) a framework by combining the individual pieces.

There is ample evidence in the literature that both kind of processes and types of knowledge structures occur. Clearly, people have many expectancies which guide their understanding and problem solving and yet they cannot possibly have an expectancy for everything. Of interest to us then, is the role of both interpretative and generative processes in expert problem solving. In particular, we are interested in such things as: what features of problems might predispose an expert toward one type of knowledge use and processes versus the other? Does the type of job or task the expert have influence his/her processing? What kinds of knowledge structures are used in each of these modes of processing and what role do these knowledge structures play?

One hypothesis (perhaps a rather simpleminded one) regarding the relationship between problem features and interpretative versus generative processing is that to the extent to which problems are familiar and stereotypic, experts will employ interpretative processes, while to the extent to which problems are novel or atypical, experts will employ generative processes.

The question of whether problem solving modes can be matched to jobs is an interesting one since it has implications for which experts to pick for which jobs. For example, if some experts tend to be better at interpretative processing than generative, they may be better suited for jobs which require such processing. We hypothesize that jobs which by nature are somewhat ill-defined (e.g. in terms of procedures to follow or products to produce) will tend to be more generative in nature, since structure is not being imposed on the expert, thus leaving him or her to build the structure him/herself. Also, jobs which require high-level integration of several bodies of knowledge or databases are hypothesized to be more generative for the following reason. Research in the cognitive bias literature indicates that people have a great deal of trouble integrating different bodies of knowledge, suggesting that they do not have well developed knowledge structures for doing so and hence are left to improvise. A more theoretical argument for this hypothesis is that the more bodies of knowledge that are required to be integrated, the greater the number of permutations which are possible for integrating those separate bodies. Hence, it would be more difficult to fit all of those potential permutations into a single framework. Even if it were possible, it is unlikely that the expert would have had enough experience with each of the possible permutations (or even a significant fraction of them) to have build a well-formed all-encompassing structure (although one measure of expertise might very well be the degree of development of that integrative structure).

Conversely, jobs with very well defined procedures are hypothesized to be more interpretative in nature. Similarly, jobs involving routine tasks are hypothesized to be more interpretative since the person performing them can develop standardized procedures for doing them. Finally, jobs which are highly goal-directed are hypothesized to be more interpretative in nature since these goals set the context for future knowledge use.

Given the distinction between interpretative and generative approaches to structuring knowledge and information processing, we examine what types of knowledge structures might support both interpretative and generative frameworks. The most common knowledge structures associated with an interpretative framework are scripts, frames, and other types of schema. While a variety of generative structures also exist, the ones that receive most support as being valid from both psychological and AI standpoints are production rules and semantic networks. Recently, much interest has been generated in the notion of mental models, schemes which model and simulate physical events. We hesitate to characterize mental models as either interpretative or generative structures since their flexibility in what they can model suggests that they can be used in either top-down or bottom-up processing.

Rather than selecting from among these various structures and using single structures as models of top-down or bottom-up processing, we believe that experts represent knowledge using a variety of structures and, therefore, argue that the most psychologically valid representation of an expert's knowledge will utilize a variety of knowledge structures. Our interpretative approach to modelling expert knowledge uses scripts, frames and mental models, and our generative approach uses production rules, semantic networks and mental models. We give a brief discussion of each of these structures and show how they are integrated in both interpretative and generative approaches.

Our view of scripts is based on the framework developed by Schank and Abelson (1977) and the Yale AI Lab. A script is a stereotypic sequence of actions, typically associated with particular physical contexts, shared across people and designed to achieve a particular set of goals. Scripts contain information regarding those goals, different variations in the way scripts are instantiated (called "tracks"), a set of circumstances which initiate the script (called "entry conditions"), a set of roles or participants that occur in the script, a set of props that are found in the script, a list of outcomes that occur when the script is executed, and a list of scenes which contain actions to be followed.

While many variations of frames have been used (in fact, one could call scripts a type of frame), frames in our framework serve as organizers of objects (something that scripts do not do well). Examples of frames include "motorized rifle division", "collection management and dissemination section (CM&D)" (the physical setting, not the procedures carried out there), and "deliberate defense" (the arrangement of units as opposed to how the battle is fought). Frames, therefore, contain slots for describing the makeup and interrelationships of the objects within the frames. Such slots include the header, what higher level frame the current frame is an instantiation of, the component parts, how they are configured, what context the frame occurs in, what other frames the current frame is similar to, physical descriptors of the frame, and an attribute list to describe the characteristics of the frame and its objects.

The representation of mental models presents an interesting challenge to the cognitive science community. To the best of our

knowledge there is no truly well-defined definition of what they are, let alone a "standard" way of representing them. For the purpose of the framework that we are proposing here, we view mental models as simulations of the causal mechanisms that underlie events. This makes an important contribution to our framework since representation schemes such as scripts and production rules may have implicit causal relationships within them, but lack an explicit representation of why the actions prescribed are prescribed or how they serve the overall objectives implied in the knowledge structure. Mental models fill this important gap.

Given that we view mental models as representations of causal mechanisms, we have constructed a representation template for mental models. This template contains five slots. The first slot contains contextual information relating to conditions or settings that the model describes (such contexts could range from being very situationally specific to being general across all situations). A second slot contains information concerning the objects that are relevant to the mental model. These objects can be both animate (such as the actors in a given situation) and inanimate. A third slot contains information on the forces that operate in the context. By forces we mean those "things" which can act upon objects to produce events. Forces can be physical such as gravity, electricity or a push and can also be more intangible such as love, hunger or determination. The fourth slot is the heart of the mental model, describing how the forces interact with the objects. Finally, the fifth slot contains outcomes that are produced when the events depicted in the model are executed.

As for generative structures, production rules have the standard representational format found in the literature (cf. Newell & Simon, 1972). Production rules have the form "If [antecedent], then [consequent]", where antecedents refer to test conditions that if met trigger the consequents, and where consequents are procedures to be executed.

Production rules can be combined in two ways. First, production rules can be chained so that once consequent procedures have been executed, they become the antecedent conditions for the next production rule. For example, a production rule sequence for a river crossing might be "If unit comes to river, then have engineers build bridges. If bridges have been built, then have troops cross bridge, etc." Second, production rules can be generalized so that either the same consequent procedures could be applied to antecedent conditions that form a generalized hierarchy, or individual procedures which apply to a set of antecedent conditions may be generalized. An example of the first case might be "If a brigade command post (CP) is seen communicating with a division CP, then infer that the brigade is subordinate to the division" leading to a more generalized rule such as "If a unit is seen communicating with a next higher echelon, then infer that the lower unit is subordinate to the higher unit". An example of the second case might be "If unit is in a defensive posture, then move artilley to the rear", "If unit is in a defensive posture, then move command post to the rear", etc. leading to a generalized procedure of "If unit is in a defensive posture, then move assets to the rear".

Finally, semantic nets are used in our framework in basically the same way they are used elsewhere in the literature. Semantic networks contain nodes and arcs (or links) that connect nodes. Nodes are specific concepts or objects and the arcs describe relationships between concepts. Sample arcs include "isa" (an object is an example of a more general category, e.g. Abrams "isa" tank), "part-of" (an object is a part of some set or larger entity, e.g. a battalion is "part-of" a brigade), and "has-part" (the features of an object, e.g. a jeep "has-part" wheel).

Note that semantic networks and frames are very similar since both describe relationships between objects. We distinguish between the two in that semantic networks seem to organize knowledge about individual objects and knowledge of other objects brought to bear in so far as it expounds on what is known about the individual object, whereas frames seem to organize knowledge about collections of knowledge and include more elaborate relationships among objects (such as physical configuration).

In order to understand how the different types of knowledge structures outlined above get integrated into interpretative and generative representation frameworks, we outline what we see as the role each knowledge structure plays in terms of the knowledge it best captures. In the interpretative framework, we view scripts as serving three main functions: capturing the general structure of the domain (in terms of the context, players, scenes, etc.), capturing the goal-based knowledge organizing principles, and capturing the general procedures that occur in the domain. Frames serve to represent integrated or high-level objects (e.g. "division", "offensive posture"). Finally, mental models show the causal reasoning that underlies procedures and links goals and actions by showing how these actions causally help achieve goals.

Given the respective roles these knowledge structures play, scripts form the heart of our integrated interpretative representation. Concepts referred to in the script (e.g. in the props or entry conditions) have pointers to frames which elaborate on the knowledge contained in these concepts. For example, a script for situation development might contain as a prop, the all-source production section (ASPS). Hence, the ASPS would be listed as a prop in the script with a pointer to a frame which describes the ASPS. Finally, mental models would be linked to script scenes giving the relationship between the script scene and the script goals,

e.g. the situation development script may have a scene "request intelligence from CM&D" which would point to a mental model showing the relationship between collecting information on the enemy's movements composition, etc. and the goal of understanding what the enemy can do to oppose the friendly unit's mission. In addition, mental models would also be linked to actions within a scene to explain their relationship to the scene's objectives.

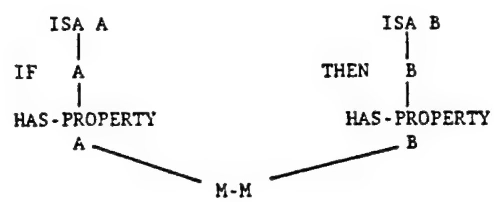
A general schematic for the integrated interpretative representation scheme is shown below:

```
Script:   A
Track:    A1
Roles:    RoleA, RoleB, etc.
Props:    PropA (pointer to PropA frame),
          PropB (pointer to PropB frame), etc.
Entry
Conditions:    ConditionA (pointer to frame
               giving attributes of ConditionA),
               ConditionB (pointer to frame
               giving attributes of ConditionB),
               etc.
Goals:    GoalA, GoalB, etc.
Results:  ResultA, ResultB, etc.
Scene 1:  header (pointer to M-M)
          Action1a
          Action1b
          etc.
Scene 2:  header (pointer to M-M)
          Action2a
          Action2b
          etc.
Etc.
```

In our integrated generative representation framework, production rules are viewed as capturing problem solving procedures. Semantic networks are viewed as capturing concept and object related information. Finally, mental models give the underlying causal reasoning for why procedures are carried out (e.g., why a unit will move its artillery forward if it's on the offensive) and serve to link semantic information to production rules (very often production rules will be dictated by the features of the objects referred to, e.g., the reason why tanks are not used in heavily wooded areas depends on the properties of tanks--their size, maneuverability, etc. Mental models can explain the relationships between these features and the procedures they influence.).

As a result of the roles that these knowledge structures play, production rules form the heart of our integrated generative representation scheme. Semantic information is linked to concepts used in the production rules and also serves to generalize or specify production rules [e.g. a rule that says autobahns are high speed avenues of approach can be specified (generalized) to say that a particular autobahn (any concrete highway) is a high speed avenue of approach]. Finally, mental models are linked to either production rules themselves or semantic features depending upon their level of explanation. For example, a mental model explaining why forces are massed on the attack may be linked to the rule itself and not any particular feature of "attack" or "forces", while a mental model explaining why tanks should be maneuvered in open ground may be linked to features of tanks.

A general schematic for the integrated generative representation scheme is presented below:



Given our two representation schemes for modeling expert knowledge, we turn to the issue of selecting knowledge elicitation techniques based on these schemes. Our view is that elicitation techniques should be selected so as to fill the structures we are using as our model of the expert, while at the same time be sufficiently flexible to allow for updating or expanding our model, if necessary.

In selecting elicitation techniques to build an interpretative model of the expert, we are interested in techniques which we feel will capture general goal and plan-based knowledge, get at the control mechanisms underlying problem solving, get at organization of concept knowledge, and finally get at the causal underpinnings of the expert's knowledge. To accomplish these goals, we select a variety of techniques. Our two major "interpretative" techniques are a structured script based-interview, with questions directed at generating scripts based on the structure outlined above, and think-aloud, "top-down" problem solving, which involves having the expert work on a standard problem and focus on his expectancies, hypotheses, etc. In addition, experts are asked to focus strongly on the rationales behind what they are doing. The interview is designed to get at the general structure of the script and any associated frames. The think-aloud problem solving is designed to fill in the specific actions of the script, get at problem solving control mechanisms, and elucidate the causal reasoning used in generating mental models. These two major techniques are supplemented with three additional techniques. One is "what-if?" extrapolations which can be used in the context of problem solving. Here, the expert is asked to imagine variations on a situation and to say how this would affect the problem. This technique is designed to get at causal reasoning processes used in constructing mental models. Another technique is to ask the expert to list

"critical instances", i.e. real-life examples of particularly good or bad solutions to a problem. This technique is designed to get at the rationale behind good and bad problem solving methods (useful for constructing mental models) and also to get at important memory organization features (e.g. what are the important dimensions along which experts process and store events). A third technique involves asking the expert to give examples of scenarios similar to the one he/she is working on. This technique also is designed to get at features of events which serve as memory organizers.

Below are sample questions for each of these techniques.

Technique Questions

Structured interview technique:
"What are your goals or objectives here?"
"Whom do you interact with?"
"Describe, in general, the procedures you go through."

"Top-Down" problem solving technique:
"What information are you looking for?"
"What hypotheses do you have?"

"What if" extrapolations technique:
"What if we change the problem to [...] how would this affect you?"
"What events would make things really good for you in the scenario, and why?"
"What events would make things really bad for you in the scenario, and why?"

Critical instances technique:

"Describe a particularly good or bad instance of [OB/situation development]"

Similar instances technique:
"Is this problem similar to others you've worked and if so, how?"

In selecting techniques to build a generative model of the expert, we are interested in techniques which will capture problem solving procedures, get at how experts handle individual pieces of data and integrate them into an evolving picture, elucidate the general concept knowledge they bring to bear, and illuminate the causal reasoning which underlies the expert's solution to his/her problem.

Two primary elicitation techniques are chosen to achieve these objectives. One technique we term "data-oriented problem solving". This technique involves having the expert work on a problem in the absence of the usual background and contextual information associated with the problem. Rather, the expert is fed individual pieces of data from a common problem and asked to respond to them in terms of what they mean to him/her, what inferences can be made from them, what further inferences could be made if additional information was given, how the current information fits in with previous information, what the problem as a whole looks like after each data point, etc. When the elicitor's supply of data is exhausted, the expert is asked to request the one datum he/she would most like to have and the elicitor provides it (even if he/she has to make it up--if he/she can't, then the expert can make up the datum).

This technique is designed to have the expert solve a problem in a truly "generative" fashion, i.e. build up a solution from bits of data, limiting the amount of context or expectancies he/she is given (although one can never eliminate expectancies entirely). Such a technique is designed to elicit information regarding production rules and semantic information associated with individual concepts and conditions and to see how pieces of information get linked.

This technique is supplemented with a structured interview, done in the context of the problem solving, which is designed to elaborate on and fill in gaps associated with the data-driven problem solving technique. Again, experts are asked for underlying causal reasoning (to help build mental models), to give generalizations or more specific instantiations of rules, and to answer hypothetical questions about variations in the situation. These techniques can also be supplemented with critical incident and similar incident techniques.

Below are sample questions for the two major generative elicitation techniques:

Generative Questions

Data-oriented problem solving:

"What does this piece of data mean to you?"
"What additional information would you need to make more sense of this?"
"How do these data fit with previous data you have received?"
"If you could have one piece of data now, what would you want to know?"

Structured interview:

"Is this a general rule?"
"Under what circumstances would you use this rule?"
"Why do you use this rule/how does it work?"

METHOD

Overview. The study reported here looks at two factors: elictitation method and task type. Both factors were between subjects. The elicitation methods used were the interpretative and generative methods outlined above. Subjects were interviewed in the context of two constrasting tasks to test the robustness of each elicitation method and to examine what charac-

teristics of expertise might be best captured by what methods/techniques. The tasks chosen were situation development (SD) and order of battle (OB), both intelligence tasks performed in the G-2 shops.

Situation development was viewed as a high level task, requiring the integration of several sources of knowledge in order to build a picture of the enemy situation, thus having certain generative characteristics. On the other hand, the situation developer deals directly with goal and mission-related information (i.e. his own unit's mission) which strongly influences his evaluation of the situation. Hence, situation development has an interpretative component as well.

While OB has a great deal of similarity to SD (both are concerned with the enemy situation), there are some interesting contrasts as well. The OB technician or analyst does not have to integrate as many different bodies of knowledge as the situation developer and is not charged with constructing "the big picture". Hence, in this respect, OB is not as generative as SD. Rather, the OB technician or analyst has a variety of templates regarding the enemy's order of battle onto which he maps incoming information. Hence, the OB technician's job is highly interpretative--in fact, using our framework, we would say it is highly frame-oriented. On the other hand, OB technicians typically deal with data at a much lower and more detailed level than situation developers, hence making them more generative in this regard.

Subjects. Subjects were recruited from three Army installations in the continental United States: the G-2 shops in the Divisions at Fort Bragg and Fort Carson, and at the Intelligence Center and School at Ft. Huachuca. A total of 20 subjects participated in the study. Of these 20, ten were interviewed as experts in order of battle and ten as experts in situation development. Determination of area expertise was done by those representatives of each installation who were in charge of scheduling interviews. Half the subjects in each task were interviewed using the interpretative elicitation method and half were interviewed using the generative elicitation method. The amount of time spent with each subject ranged from about three to ten hours, depending upon the subject's schedule.

Materials. Two basic types of materials were used in the study. One was a standard (Fort Leavenworth) training problem, in which a U.S. division, forming part of a U.S. corps in a defensive posture against a Soviet Combined Arms Army (CAA), is ordered to counterattack to achieve a specific objective. The materials included in the problem were a map of the relevant terrain, overlays depicting the current known enemy situation, enemy order of battle templates, a corps periodic intelligence report (perintrep) and an intelligence summary (intsum) put out by one of the brigades that were in contact with the enemy (through whose sector the subject's division would be passing). The perintrep and the intsum represented incoming information that the subject would use to update the situation and build whatever picture he was building.

The second type of materials used were data recording sheets which contained templates for each of the knowledge structures associated with the elicitation method being used with the subject. These coding sheets allowed the experimenters to keep track of the kinds of things the subjects were saying and guided the experimenters' questions by showing them what slots in the knowledge structures still needed to be filled.

Procedure. All subjects received an initial overview of the project's purpose, namely to explore methods of eliciting and representing expert knowledge. Subjects were assured that the experimenter's methods (not their performance) were being evaluated and that all data collected would be reported anonymously. The elicitation teams contained two primary members (although there were often additional observers): an interviewer, and a retired general officer with expertise in the intelligence domain. The general officer answered questions about our procedures from a military point of view, and also played the roles of people with whom the subject would normally interact when working on actual problems. All interviews were tape recorded and later transcribed. These transcripts, along with notes taken during the elicitation sessions, served as the source of data.

The above procedures were common to subjects in both the interpretative and generative elicitation method conditions. The rest of the procedures were different and are described below. In the interpretative condition, subjects began with a general interview concerning the relevant area of expertise. This interview was designed to get general information and set the context of the remainder of the session. Following the general discussion, the experimenter asked a series of script-relevant questions (guided by the script-based data sheet). This was followed by having the subject work on the problem, while thinking aloud and describing his thought processes. Finally, if time permitted, the supplemental techniques (e.g. critical incidents) were employed. Throughout the session, the experimenters asked a series of questions to get the subjects to elaborate on or explain their thinking. The subjects were often challenged by the experimenters and asked to explore a number of alternative solutions.

The generative condition focused primarily on problem solving. Here subjects began receiving individual pieces of data with no background information other than what was contained in each datum or what they could infer from it.

This procedure continued until subjects could make no further progress without additional information, such as a map or knowledge of what their unit was or their mission, etc. Only then was such information provided. Throughout this procedure subjects were asked probing questions to uncover their thought processes, what inferences and hypotheses they were making, etc. Note that in this condition, since subjects were attempting to generate their own context, they often made gross mistakes in terms of the units they were facing, the location of the FEBA, and whether the enemy was on the offensive or the defensive. In a realistic situation of course, the subjects would have had much more information. What was interesting was to see how subjects updated their conclusions and modified erroneous assumptions. If time permitted, subjects were exposed to supplemental elicitation procedures such as critical incidents.

RESULTS AND DISCUSSION

In the present study, we are interested in two types of results. One is what are the relative strengths and weaknesses of the interpretative and generative approaches to knowledge elicitation in terms of the knowledge they can best capture. The second is what might be the features of tasks/situations which which might make either an interpretative or generative approach to knowledge elicitation more effective.

Based on our sessions with the experts, we have reached the following conclusions regarding the strengths and weaknesses of the interpretative and generative methods of knowledge elicitation (as employed in our study given our time constraints with each subject). First, the interpretative method is very good at eliciting goal and mission-based processing. It is also good at giving the overall structure of the expert's problem solving approach. Its weaknesses lie in eliciting knowledge regarding concrete details (experts would often gloss over these), and the background assumptions and knowledge that the expert is using. Experts have a well of knowledge they bring into a problem which is triggered by the context they are given. It is virtually impossible, using this method, to get them to state all their assumptions (although they do state many of the major ones).

Interestingly (and fortunately), the strengths and weaknesses of the generative method appear to be complementary to those of the interpretative method. The generative method is good at eliciting the detail knowledge utilized to handle individual pieces of data. It is also good at eliciting how experts integrate data. This method also offers a partial solution to the "assumption problem" described in the previous paragraph, since many of the implicit assumptions are removed and the expert must be more explicit in the assumptions he is making in order to determine the implications of each piece of data. The weaknesses of the generative method are that it is not very good at eliciting goal-related information (but this may be due in part to time constraints not permitting the experts to work at a more "top-level" analysis where overall problem-solving goals, such as the unit's mission might play a more prominant role) and that it is not particularly good at eliciting problem solving control mechanisms. This latter weakness may be an artifact of our methodology since we controlled the data that the subjects received and the order in which they received it. This may have interfered with how they would normally solve the problem.

We now turn to the second of our major areas of interest in this study, namely features of tasks or situations which might be more conducive to interpretative or generative approaches to knowledge elicitation. The first feature we have identified is whether the task seems to call for more procedural or content knowledge. In our study, situation developers appeared to be more procedurally oriented, while order of battle technicians appeared to be more content oriented (in fact, they were often described as "walking encyclopedias").

Content knowledge seemed to be highly frame or semantic-based and seemed to respond better to generative or data-driven techniques. Procedural knowledge seemed to revolve more heavily around scripts, production rules, and mental models and responded better to interpretative or theory-driven techniques.

A second feature was whether the task can be described as "high-level" or "low-level". In our study, high-level refers to the more large-scale integration of bodies of knowledge by situation developers, whereas "low-level" refers to the more small-scale pattern recognition or template matching that order of battle analysts do. These pattern recognition processes seem to respond better to "interpretative" techniques (such as direct slot-driven interview) whereas the high level integration processes seem to respond more to "generative" techniques (such as data-based problem solving).

One qualification to our arguments that knowledge tends to be represented in specific kinds of structures and hence can be best elicited using specific techniques is the finding that the same knowledge can often be represented in different formats (or perhaps get transformed into different formats as they are needed, depending on the context the expert is in) and hence be elicited using different methods. Below is a frame of a Motorized Rifle Division Defense, observed by the interpretative method:

Motorized Rifle Division Defense Frame

Frontage:	20 - 30 km
Composition:	Pointer to motorized rifle division. May have assets attached from Army or front
Configuration:	2 MRR on line, 1 TR in reserve Division CP to the rear Artillery to the rear (2/3 range behind FEBA) Air defense to the rear (2/3 range behind FEBA) General Rule: Assets are placed to the rear
Explanation:	Defender will be retreating while attacker will be advancing. Placing assets to rear keeps them less vulnerable to attack while targeting attacker as he moves forward.

Such a frame can be used to derive production rules useful in making inferences regarding the enemy situation, as shown below:

Inference Rules Derived From Defense Frame

If see artillery starting to be moved back, then enemy unit is likely to be planning to be on the defensive

If the forward air defense is starting to be moved back, then enemy unit is likely to be planning to be on the defensive

General rule: If assets are starting to be moved back, then enemy unit is likely to be planning to be on the defensive

---

These inferences were elicited using the generative method while experts were working on a problem. This example shows how the same knowledge can be represented differently, depending upon what task (including how the question is asked) the expert is faced with. The implication for cognitive scientists, knowledge engineers, etc. is that knowledge may not have an "absolute" representation, but may adapt to the task the expert is faced with. In fact, this may be one of the important characteristics that define what makes someone an expert, namely the ability to transform his/her knowledge to a form most suitable for use in a particular task.

With this last comment, we would like to turn to a final issue which arose in the course of our study, namely what makes someone an "expert". We found (as undoubtedly other knowledge elicitors have found) that some of our experts seemed to be much more "expert" than others. We, therefore, tried to generate a list of criteria as to what we thought distinguished experts from non-experts.

Experts, we feel, have a great deal of content knowledge regarding their domain. This was evidenced in our study by subjects who seemed to know a great deal about things like Soviet doctrine and equipment versus those who continually had to refer to manuals for such information. If such information is already possessed by the subject, it is reasonable to expect that such knowledge would be linked to other knowledge and hence more available for problem solving.

Experts are better at recognizing/generating possible interpretations or problem solving solutions and are capable of testing more alternative hypotheses. As a result, experts are better able to develop innovative problem solving procedures (often based on experience) and know when to deviate from standard or doctrinal solutions.

Experts are better able to integrate events and form conclusions based on such integrations. Related to this, experts are better able to reason at different levels of abstraction. Subjects whom we felt to be less expert seemed to have trouble seeing "the big picture" and reason at high levels of abstraction.
Finally, experts have a good understanding of the reasoning that underlies their problem solving procedures, or to use our termninology, have well-articulated mental models. We found that by understanding why they were doing what they were doing, "expert" subjects were better able to evaluate what procedures were appropriate and useful and to generate new ones, if necessary.

Based on our framework and our findings, we have identified three application areas that our work could be useful in: identification of experts relative to particular domains or jobs, training of students to acquire expertise, and the design of expert systems and decision aids.

The problem of identifying experts is particularly important when deciding who is going to work what job or be relied on in a decision-making role. In addition to the general characteristics of experts which we outlined above, we have also been exploring the characteristics of tasks and jobs and the demands they place on the kind of knowledge needed to do them. Our work may eventually allow us to analyze the knowledge requirements of a particular job or task and then match an expert to that task.

A related issue is the training of experts. If we have an understanding of what characteristics define an expert and how that expert uses his/her knowledge, training programs can be developed which are designed to foster that kind of knowledge and problem solving. Similarly, the characteristics of the particular job the student is training for could influence what is taught, e.g., rote (or content) learning versus

more practical (or procedural) learning. Also, the elicitation methodology itself could be adapted as an evaluation tool to determine the effectiveness of the training program, i.e., to help teachers evaluate the level of the students' expertise, including gaps in the students' knowledge.

Another application of our framework would be in the selection of representation schemes for expert systems, based on the characteristics of the domain the system is modelling. Expert systems which are modelled upon the way experts represent and use knowledge will be better able to aid experts in their jobs. Such systems would both be more "expert" and more compatible with the expert using it.

REFERENCES

1. Newell, A. and Simon, H.A. (1972) Human problem solving. Englewood Cliffs, N.J.: Prentice-Hall.

2. Schank, R.C. and Abelson, R.P. (1977) Scripts, Plans, Goals and Understanding. Hillsdale, N.J.: Erlbaum.

---

*This research is sponsored by the Army Research Institute under Contract No. MDA903-86-C-0383. The COTR is Dr. George Lawton.

Working Paper

WP HUA 90-03

An Evaluation of Tests of Sensitivity to Nonverbal Communication

Philip H. Marshall, Texas Tech University, Lubbock, TX

August 1990

Reviewed by: JULIE A. HOPSON
Chief, ARI Field
Unit-Ft. Huachuca

Approved by: JULIE A. HOPSON
Chief, ARI Field
Unit-Ft. Huachuca

Cleared by: ROBIN L. KEESEE
Director, Systems Research
Laboratory

U.S. Army Research Institute
for the Behavioral and Social Sciences
5001 Eisenhower Avenue, Alexandria VA 22333

AN EVALUATION OF TESTS OF THE SENSITIVITY TO NONVERBAL COMMUNICATION

CONTENTS

List of Figures

AN EVALUATION OF TESTS OF THE SENSITIVITY
TO NONVERBAL COMMUNICATION

Introduction

A recent study on Characteristics of Successful Military Intelligence Interrogators (Knapp, 1989) was conducted to ascertain those characteristics that could be evaluated and used to refine screening procedures of future 97E10 candidates. Through a series of filtering procedures, eight factors were finally determined. These included: (1) good foreign language skills; (2) common sense; (3) well rounded background, intellectual knowledge; (4) flexible, adaptable to any situation; (5) knows military tactics, equipment, organizations; (6) keeps control of situation at all times; (7) communicates easily with people, and (8) picks up subtle cues, nonverbals.

Two existing psychological inventories have been selected to evaluate factors 1 through 7. Those tests are the California Personality Inventory (1987) and the Myers-Briggs Type Indicator (1988). The purpose of this study was to determine an appropriate instrument for the last factor, "picks up on subtle cues, nonverbals."

Social scientists have long recognized the importance of nonverbal cues in affecting interpersonal behavior, and that "accurate social perception depends on the ability to observe and interpret the expressive behaviors of others" (Costanzo & Archer, 1989, p. 225). Several decades of research have resulted in a number of tests for the ability to interpret nonverbal communications. These tests have taken several approaches, and some have met with more success than others.

The Brief Affective Recognition Test (BART) (Ekman & Friesen, 1974) requires the brief presentation of slides of facial expressions that subjects have to correctly classify for emotional content. The Communication of Affective Receiving Ability Test (Buck, 1976) asks the subject to identify what type of scene the subject in the slide is viewing.

To date, the most widely used test has been the Profile of Nonverbal Sensitivity (PONS) (Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D., 1979). This test has the subject view over 200 videotaped segments of a single female subject demonstrating a variety of affective behaviors. The subject's task is to identify what the person on the tape is doing. Costanzo and Archer (1989) have listed several "constraints" associated with the PONS. These include the "reliance on a single encoder, the possibility that subjects might recall cues from one scene to another, presentation of

communication channels in isolation, use of posed emotions and situations, and use of a problematic criterion of accuracy--the affective intent of the encoder" (p. 227). There is the additional criticism that the dress and appearance of the encoder (the person on the tape) has become out-of-date in light of the 20 years that have passed since the instrument was first devised.

Costanzo and Archer (1989) have offered a new test (the Interpersonal Perception Task) that has some of the same objectives as the PONS, and is considered an improvement. This test presents the viewer with 30, short, videotaped segments of individuals interacting in natural situations. The people in the segments are not actors, and their interactions are unscripted. They include persons of both gender, different racial heritages, and an age range from about two years to "retirement." The 30 segments (each of which lasts from 30 to 60 seconds) are divided into five categories of interpersonal relationship (kinship, deception, competition, status, and intimacy). The subject is first shown the question to be answered (e.g., What is the relationship between the man and the woman?, Which person is telling the lie?), sees the actual segment, and is given several seconds to select the correct alternative on the answer sheet. For each segment there is a single objectively correct response.

In their review and validation of the test, Costanzo and Archer (1989) report that in their normative sample of 438 college students, subjects responded above chance level for 29 out of 30 scenes, and that women had higher scores than men, not an uncommon finding on tests measuring nonverbal communication (Hall, 1984).

In order to validate their new instrument, they had a group of college women make interpersonal skill ratings on each other, and those judged to be more "socially skilled" had, in fact, obtained higher IPT scores.

Another recent instrument is the Affect Blend Test (ABT) (O'Sullivan, 1983). In contrast to the IPT, the ABT takes a micro rather than a molar approach. O'Sullivan and her colleagues base their instrument on the minute changes in facial musculature that accompany various emotional expressions. The raised eyebrow, the partially opened lips, the slight downward curl of the mouth, the openness of the eyes, and so on, have all become recognized as components of various emotional expressions. Their test consists of the presentation of 56 "facial expressions" that subjects view, and have to interpret. However, the expressions are unique in that they are actually composed of <u>blends</u> of facial characteristics. For each picture, the subject is instructed to circle one or more words that indicate what emotion is being represented. The alternatives include "happy, sad, fear, anger, surprise, disgust, and contempt." O'Sullivan's rationale for developing this test was that tests of single affects are too

easy, and single affects rarely occur in everyday life. She also reported some correlation between the ABT and an alternate form of the PONS, but not with Mehrabian's Emotional Empathy Test.

Method

The overall design of this study was to have 97E10 students and instructors take each of the tests, and then to conclude which of the instruments would be the better choice for the developing test battery.

Subjects. The students were 32 individuals (eight women) who were awaiting entry into the initial 97E10 class, or who had been in the class for only two weeks. The instructors were 21 men who were regular instructors in this course. Each was considered to be a "subject matter expert," since each had at least one active duty assignment as an interrogator.

Apparatus and Materials. A videotape player and television were needed to present the IPT. The 56 pictures on the ABT were cut out of test booklets and glued and laminated to two sides of a 15 x 15 inch mounting board (30 on one side, 26 on the other). Dr. O'Sullivan generously provided ample copies of the ABT for this purpose.

Procedure. The subjects were run in four moderate sized groups. There were two groups of students (16 in each group) and two groups of instructors (13 subjects in one, and 8 in the other). The students took the two tests in counterbalanced order, half taking the IPT first and half taking the ABT first. Since performance was not affected by test order, the instructors received their tests with IPT presented first. (This proved to be the better format since the IPT is timed, while with the ABT, test duration is subject-determined. Giving the IPT first made the testing situation a bit more manageable). Testing was complete within one hour.

Results

Performance on each test will be presented first, followed by a comparison of the two.

IPT. The score for each subject on this test was the total number of correct responses out of a maximum of 30. Note, that while the test is divided into five categories of interpersonal relationship, the authors' own analysis advises against considering these categories to represent subscales (Costanzo & Archer, 1989).

There are 15 two-alternative choice questions, and 15 three-alternative choice questions. The combined chance score was thus determined to be 12.5 (7.5 for the former plus 5 for the latter). The mean score for all the students was 15.75 (sd = 2.34), and the mean score for the instructors was 15.29 (sd = 2.12).

The distribution of scores from the combined sample of 53 subjects is shown in Figure 1. It can be seen that the scores are relatively normally distributed, with a slight negative skewness (-.12).

A *t*-test was used to determine if this difference between the two groups was significant, and it was not [$t(51)=-.732, p>.40$].

Since previous literature has indicated that women tend to out perform men on these types of tasks, another *t*-test compared the men and women students (there were no women in the instructor sample). The results of that test indicated no significant difference between them, [$t(30)=.17, p>.80$].

We also had the opportunity to compare students of another MOS with the 97E10's on the IPT. Fourteen students taking their initial advanced training course in 96B10 and 96D10 were given the IPT. The mean performance of those subjects was 13.78 (sd = 2.39), and the difference was significant [$t(44)=2.60, p<.02$].

The ST scores needed to qualify for 96B, 96D, and 97E are 105, 95, and 95 respectively. But the 97E also has to demonstrate a basic understanding of "geographic, social, economic, and political conditions of at least one foreign country," and they also have to obtain a qualifying score on the Defense Language Aptitude Battery. Thus, the 97E students may in fact possess greater all-around knowledge and interpersonal abilities.

The foregoing comparison may demonstrate that Army selection standards have been set to levels able to discriminate among abilities necessary for successful 97E10 performance. It also demonstrates that the IPT has the ability to discriminate between groups that, on the face of it, may differ.

ABT. This test is scored by taking the total number of correct emotions recognized in the total 56 pictures. Figure 2 shows the distribution of total scores on this test, and again it is relatively normal, but with some positive skewness (.56).

The students had a mean score of 48.44 (sd = 7.02) and the instructors had a mean score of 44.24 (sd = 5.68). These values were within the range expected from other studies (O'Sullivan, personal communication). A *t*-test was conducted to determine if this represented a significant difference, and it was, [$t(51)=-2.29, p<.03$]. Since the student sample included women (n=8, with a

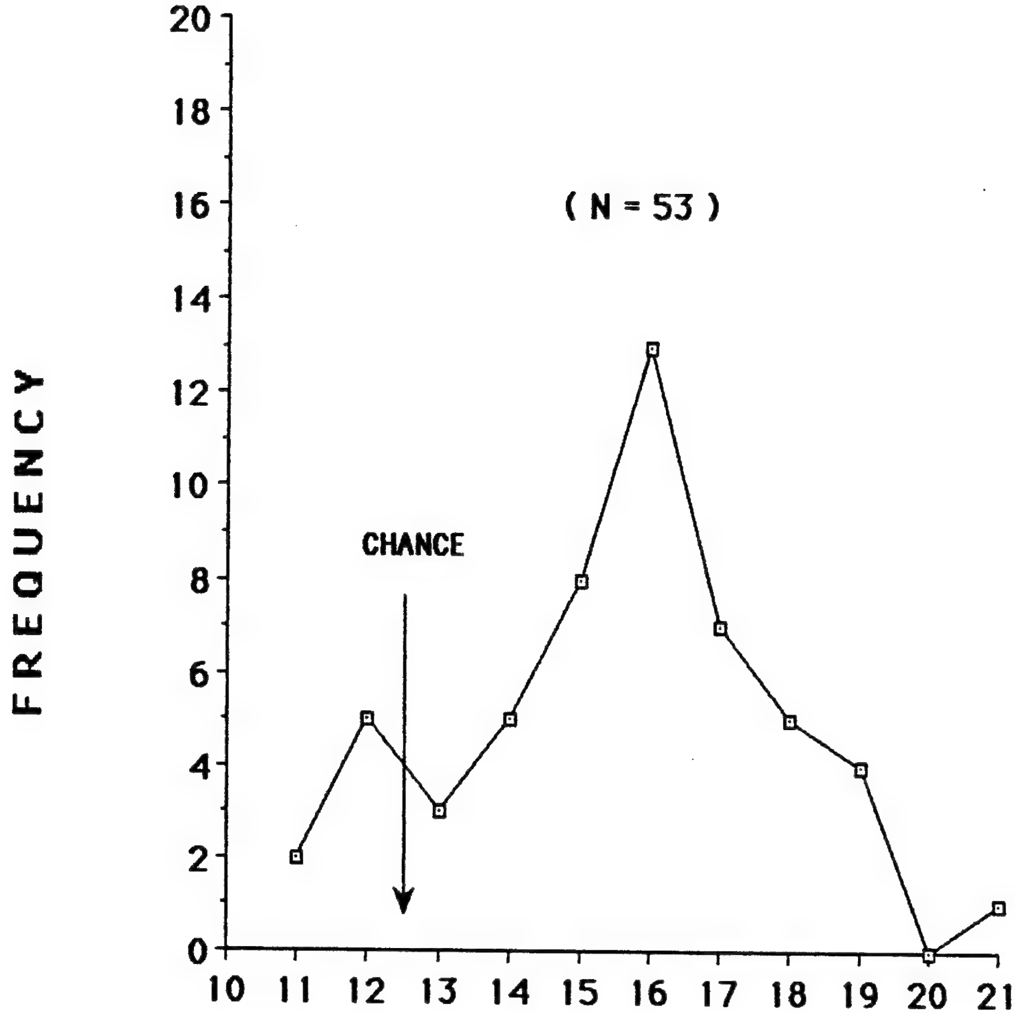


Figure 1. IPT scores.

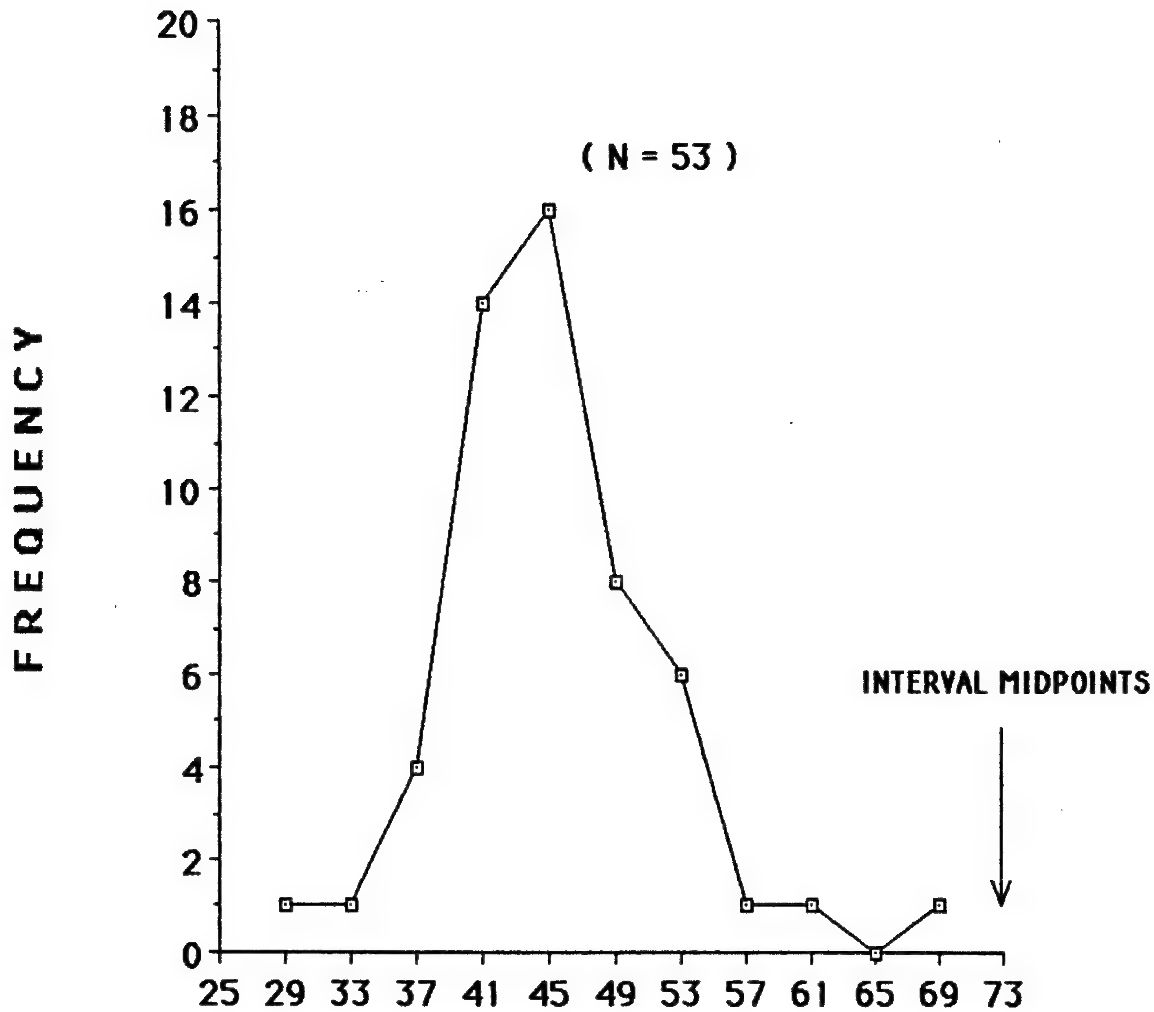


Figure 2. Affect blend test scores.

mean score of 49.88 compared to the mean men's score of 47.96) and the instructor sample did not, another *t*-test was conducted to determine if this significant difference would obtain with women removed from the student sample (i.e., if the effect was due solely to the women's superiority). The results of this *t*-test on "men only" yielded $t(43)=-1.97, p=.053$, close enough to the ".05" level to conclude that women were not exerting special influence on the difference between students and instructors.

This finding presented the interesting result of supposed experts in interrogation being less able to identify affective expressions than novice students. There is a possibility that this difference may be due to one factor other than actual skill. The instructions of the ABT allow subjects to circle one or more words that indicate what emotions they see in the pictures. The instructors may have much more confidence in their choices than the students, by virtue of their maturity and experience, and may be less likely to show what they may interpret as ambivalence by choosing more than one emotion. In fact, 29% of the instructors and only 16% of the students had "0" multiple responses. (It should also be noted at this point that no subjects were told that the photographs were actually blends of emotions, so they had no idea that multiple responses were correct.)

A *t*-test was conducted on the total number of multiple emotion selections that were made by the students and by the instructors. The mean values for the two groups were 5.14 (sd = 6.09) and 10.38 (sd = 8.61) respectively, and the resulting comparison yielded a significant effect [$t(51)=-2.41, p<.02$]. Of course, the smaller number of multiple responses by the instructors could still have been due to a lower actual sensitivity level, but this analysis offers a possible alternative explanation.

IPT and ABT. This analysis considered to what extent performance on the IPT and ABT are correlated. The correlation between the scores on the two tests for the combined sample of 53 subjects yielded $r(51)=.04, p>.78$, an insignificant value.

Discussion

The objective of this study was to make a determination of which test of sensitivity to nonverbal communication to include in the 97E10 screening battery. Several factors have to be considered, and they include the validity, reliability, psychometric properties, and ease of administration of the instruments.

The IPT seems to have the edge on validation studies. There is at least one published account of significant correlations between performance on the IPT and peer evaluations (Costanzo &

Archer, 1989). It also seems to have greater face validity in that it presents dynamic multi-channel communications. The ABT shows only a single, artificial facial expression. Thus, the IPT is more true-to-life.

In terns of reliability, O'Sullivan (1983) reports a median total score reliability of .62, while Costanzo and Archer (1989) report a test-retest reliability of .70 for the IPT. These are comparable scores and do not serve as a basis for discriminating between the two tests.

As currently presented, the ABT seems to have a problem in its standard scoring procedures. Subjects are instructed to circle one or more words that indicate what emotions they may see in the photographs. A subject who is more willing to circle a greater number of words will, by mere virtue of that, get more correct responses. As already stated, this may explain why the instructors performed less well than the students on this test.

The two tests are equally easy to present, with instructions for both being simple to follow. The ABT has the advantage of taking only about 15 minutes to complete.

The fact that performance on the two tests was not correlated indicates that they probably do not measure the same underlying abilities.

The recommendation of this researcher is to administer the IPT as part of the test battery. It offers a more real-to-life, dynamic format, it can apparently discriminate among classes of individuals who may be expected to differ in this skill (97E10's vs 96B-, 96D-10's), and it appears to have a greater degree of published validation to support it.

References

Buck, R. (1976) A test of nonverbal receiving ability: Preliminary studies. Human Communication Research, 6, 47-57.

California Personality Inventory. (1987) Palo Alto: Consulting Psychologists Press.

Costanzo, M. & Archer, D. (1989) Interpreting the expressive behavior of others: The interpersonal perception task. Journal of Nonverbal Behavior, 13, 225-245.

Ekman, P. & Friesen, W. V. (1974) Nonverbal behavior and psychopathology. In R. J. Friedman & H. M. Katz (Eds.). The psychology of depression: Contemporary theory and research. New York: John Wiley.

Hall, J. A. (1984) Nonverbal sex differences: Communication accuracy and expressive style. Baltimore: Johns Hopkins Press.

Knapp, B. (1989) Characteristics of successful military intelligence interrogators. USAICS Technical Report 89-01.

Myers-Briggs Type Indicator. (1988) Palo Alto: Consulting Psychologists Press.

O'Sullivan, M. (1983) Recognizing facial expression of emotions: The Affect Blend Test. Paper presented at the American Psychological Association, Anaheim, CA.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. (1979) Sensitivity to nonverbal communication: The PONS test. Baltimore: Johns Hopkins Press.

SRL - FT. HUACHUCA FIELD UNIT

Working Paper

HUA 88-04

A FRAMEWORK FOR TACTICAL DECEPTION

Allen L. Zaklad and Kenneth L. Moan

Analytics, Inc.

Wayne W. Zachary

CHI Systems, Inc.

Beverly G. Knapp

Army Research Institute - Ft. Huachuca Field Unit

AUGUST 1988

U.S. Army Research Institute for the Behavioral and Social Sciences

5001 Eisenhower Avenue, Alexandria VA 22333

A FRAMEWORK FOR TACTICAL DECEPTION

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In the early 1980's, Defense Science Board studies recommended that tactical deception at the Corps and Division levels be a systematic, integral part of overall operational planning, consistent with Command, Control, and Communications Countermeasures (C3CM) operations. It was also imperative to insure consistency of tactical deception initiatives between echelons Corps and below with echelons above Corps. A major step toward meeting this directive is the beginning codification of US Army doctrine on deception, now reflected in FM 90-2, Battlefield Deception. This revised doctrine - a requirements statement for the 1980's and beyond - is a necessary but not sufficient step in the evolution of the battlefield deception domain. However, before more comprehensive work is done to expand the domain, the factors or variables which are critical to tactical deception operations must be identified. This paper presents a beginning framework which supplements the FM and captures these relevant factors for consideration by those in the battlefield deception area. These apply equally to developing programs of instruction (POIs) for deception training, designing techniques to aid a deception planner, as well as for determining the necessary knowledge and materiel needed to execute a deception plan and accomplish an operational mission.

OVERVIEW OF THE FRAMEWORK

A framework is a set of interrelated structures, categories, and models that define a domain (in this case, battlefield deception). A domain is a set of individual problems or situations that may be similarly treated because they share common features that allow them to be solved in a common manner. A framework, then, defines the underlying similarity and structure of the domain as explicitly and succinctly as possible. Doing this allows those concerned with the domain to approach it at the level of this underlying structure (the "forest") instead of just specific isolated problems (the "trees"). this in turn, permits different aspects of the domain to be approached in an integrated manner. A conceptual overview of the framework structure developed for battlefield deception is shown in Figure 1. Note that this diagram consists of three large branches at the highest level, each of which contains a number of specific branches and sub-branches. Each branch will be elaborated below.

FFOR (Friendly Force) Goals Component

The first component of the tripartite framework for tactical deception defines the ends or goals that FFOR (US) deception planners may pursue. This component defines what constitutes a "solution" to an individual deception problem, because the problem will be considered solved if a plan can be devised that accomplishes the goal. The goal branch also describes the interrelationships among these possible goals and

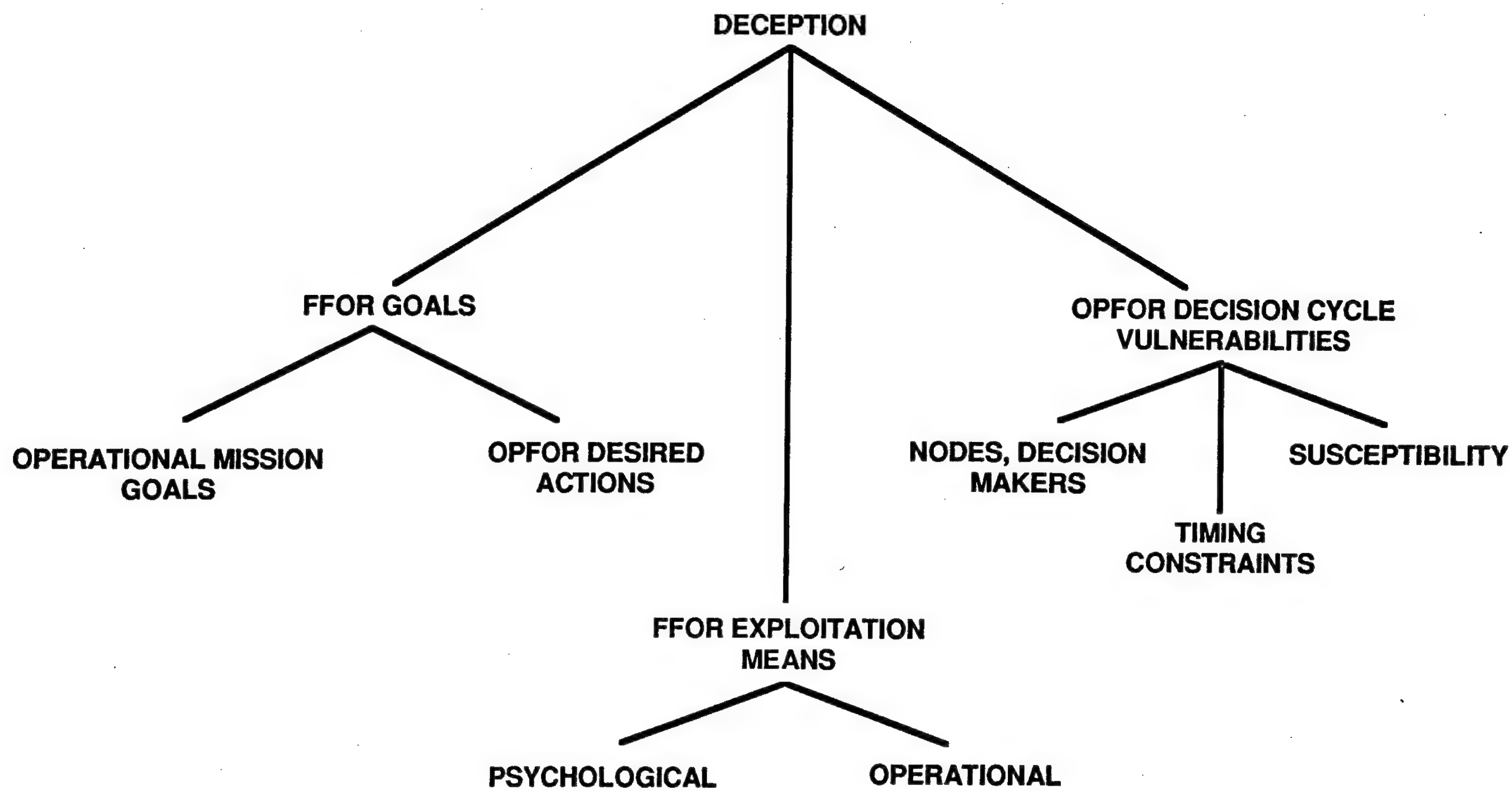


Figure 1. Overall organization of deception framework.

subgoals by placing them at various levels of analysis into a single hierarchical structure. This structure is depicted in Figure 2.

FFOR Operational Mission Goals

The first checkpoint for deception is the identification of the overall operational mission goals. By noting the commander's mission objective (e.g., attack, defend, etc.), the mission goal is established to support this mission objective. The possible goals are gain time (delay enemy action/reaction), gain enemy (OPFOR) information, or reduce OPFOR assets. Any or a combination of these three may contribute invaluably to the commander's mission objective, and subsequently drive the nature of a deception plan (what the FFOR would like to induce or reinforce in the enemy).

OPFOR Desired Actions

In concert with determination of the FFOR mission goal(s) is deduction of OPFOR desired tactical actions. This goal includes most conventional deception activities, i.e., the playing out of the desired FFOR operational goal. Under this goal, the deception planner is trying to induce or reinforce the OPFOR commander, via a deception plan, to act in a manner more advantageous to the FFOR than he would have in the absence of the FFOR deceptive measures. This deception goal supports the FFOR commander's ultimate goal of accomplishing his mission.

There are three major types or classes of tactical actions that the FFOR might like to achieve: OPFOR divert resources, OPFOR expend resources, OPFOR expose assets. Diverting and expending resources both refer to inducing the OPFOR to position or use personnel or materiel resources so that fewer resources will be available to meet a planned FFOR tactical action. Therefore the outcome distribution (spread of possible enemy courses of action) will be more favorable to FFOR. "Divert" refers to relocating OPFOR assets so that they will be out of position to respond to the FFOR action. "Expend" refers to the using up of expendable OPFOR assets, with the same bottom line result. "Expose assets" means to gain useful information about OPFOR assets based on FFOR deceptive actions. All of these desired actions will need to be considered in relation to time available for the deception, force composition, and location of enemy units.

FFOR Exploitation Means Component

The next component of the framework (cf. Figure 1) identifies the various means available to the deception planner to achieve any or all of the goals identified above. The available means are divided into two subcomponents or branches,

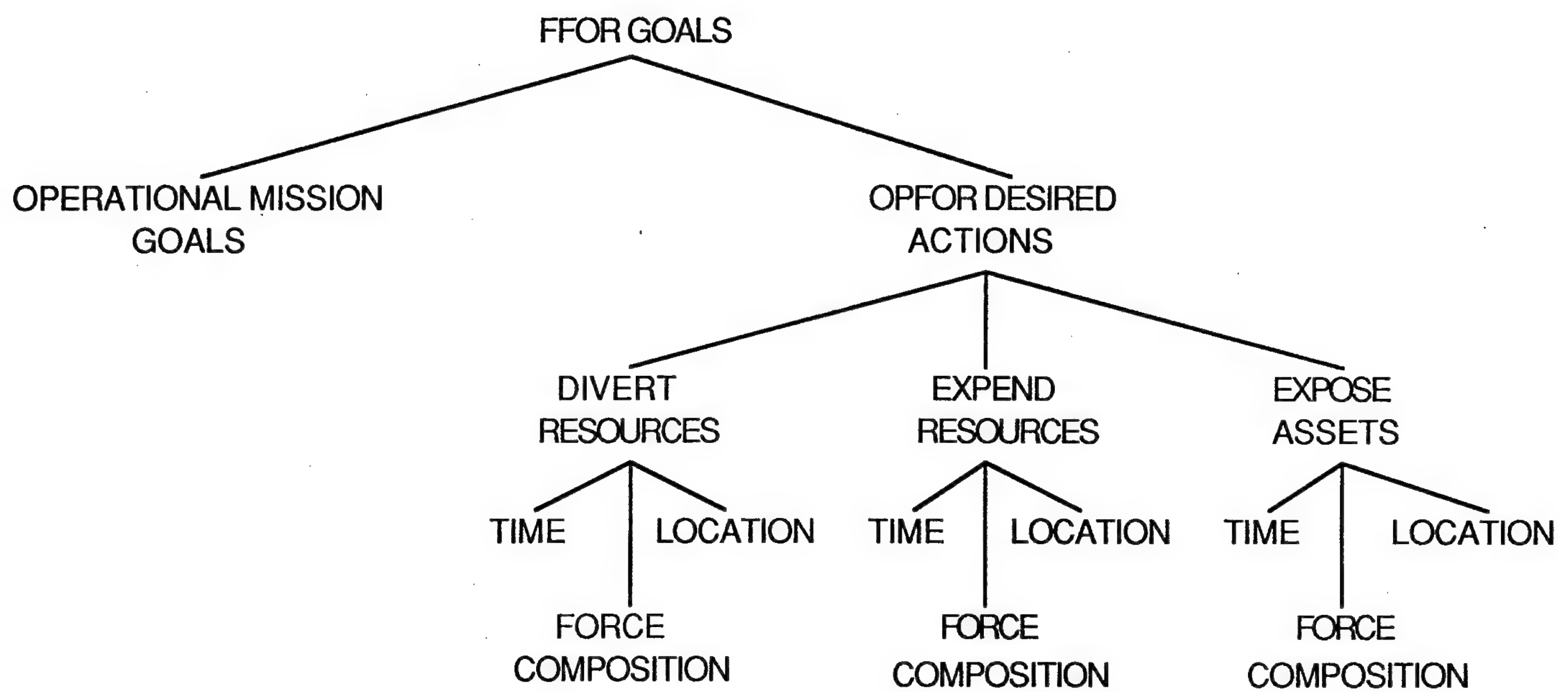


Figure 2. FFOR Goals Component

psychological and operational. Psychological means include the various ways in which the "thought processes" of an OPFOR decision maker can be systematically manipulated. However, like any tactic, they require other means to be implemented, thereby involving the other branch, operational means. The operational branch factors include the physical resources that are available to support or supplement the various operational/tactical actions and information denial/control measures. Physical resources include both standard, organic resources as well as deception specific materiel. A deception plan will involve a complex combination of psychological tactics, operational actions, and physical resources.

Psychological Means for Deception

The first branch of the deception means component refers to psychological means. Psychological means are the known avenues that the deception planner has to "get inside the OPFOR decision maker's head". In a very real sense, psychological means are the most critical means for deception. This is because tactical deception is essentially a psychological phenomenon, an attempt to manipulate the mental processes of the enemy commander. These means are shown in Figure 3. This part of the framework provides the stratagems, techniques, and approaches that need to be considered to affect enemy commander and organization, and are drawn from the areas of human information processing (perception and cognition), behaviorism (individual and group processes) and clinical psychology.

The highest level distinction in the psychological means category is that between the individual and the organizational. This refers to whether the particular technique or principle is targeted primarily to individual or to group/organizational decision making. Figure 3 indicates that tactical decision making is an individual cognitive process as well as a part of an organizational network of group processes. Individual and organizational decision making are seen to be both distinct and highly interrelated, as well as amenable to deception.

Individual psychological means. The individual segment of the psychological means structure is illustrated in the left half of Figure 3. The three categories of individual means are information processing, behavioral, and affective.

Individual information processing means. The individual information processing category is concerned with basic properties and limitations in human individual cognitive processing. All the factors in this branch deal with the human's attempts to make sense of the multitude of ambiguous and uncertain data available to him through his sensory and cognitive mechanisms. It is in a very real sense a huge data reduction problem that confronts any individual - to turn a mountain of incoming data into a manageable approximation of "what's out there" and "how things work". The psychological

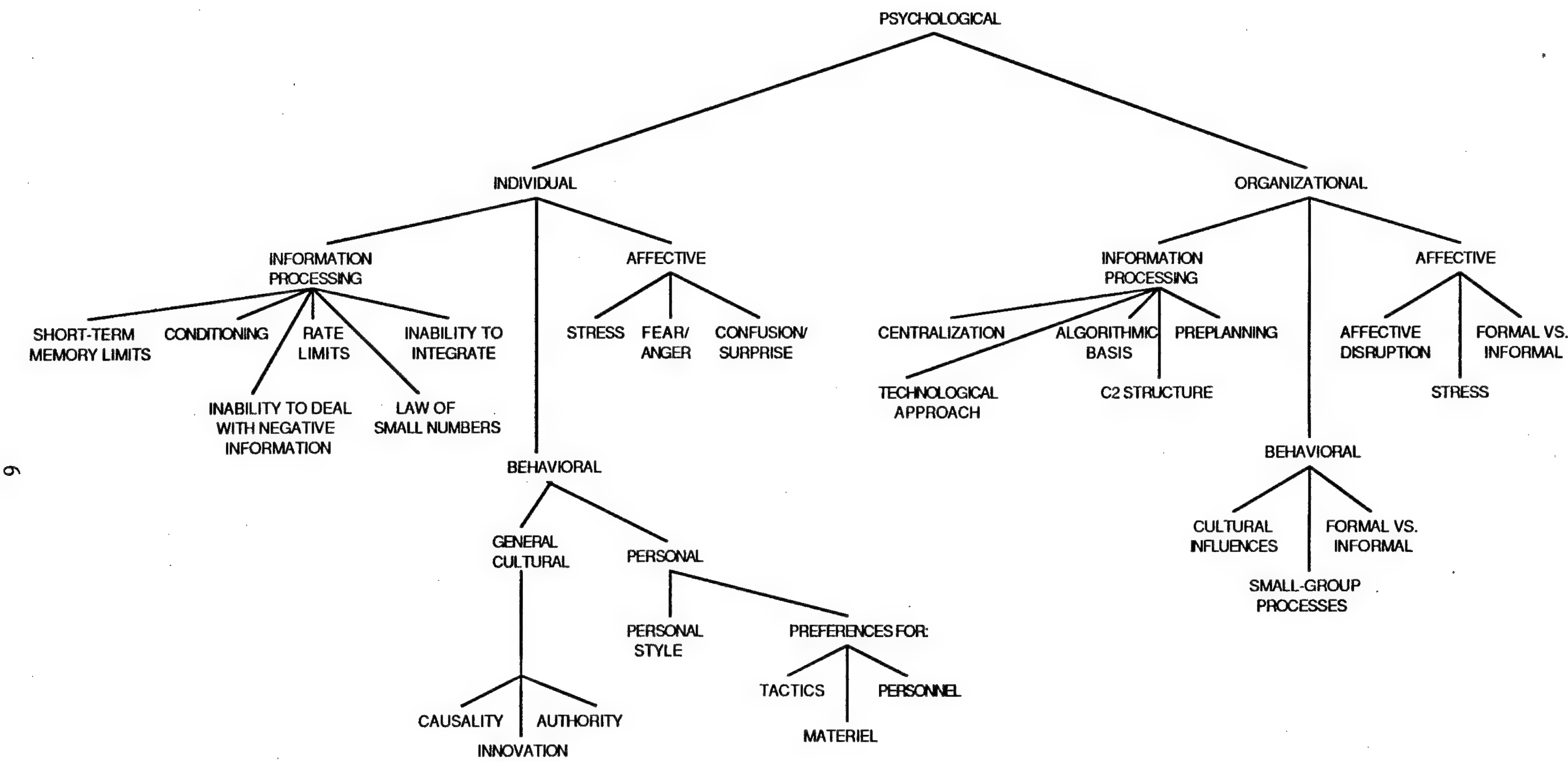


Figure 3. Psychological Means Component

deception means arising from these information processing limitations are listed and defined in Table 1.

Individual behavioral means. The behavioral means reflect points of view, perceptions, and strategies acquired during the lifetime of the decision maker. This is in contrast to the information processing means which are based on inherent cognitive characteristics. The individual behavioral means are directed toward learned or acquired cognitive knowledge and processed. These acquired cognitive tendencies consist of two classes of influences, each of which can lead to separate means of deception: cultural and personal.

Cultural influences on decision making are extremely complex and can best be viewed as providing constraints, underlying values, and preferences. An important construct in identifying cultural influences is the idea of subculture, or set of cultural influences that are unique to a specific subgroup. Each individual participates in a number of these subcultures, some with more local, immediate affect, and some with more distant, generalized affect. The various subcultural factors can be further distinguished according to both the source of cultural influence and the substantive issue involved. The disparate sources of cultural influence for the Soviet OPFOR, for example, are shown as concentric circles in Figure 4: the military, the Communist Party (all military officers must be members of the Party), Soviet culture, and finally, Russian culture.

The main substantive issues affected by the layers of culture, indicated in Figure 3, are causality, authority, and innovation. As an example, the Russian people have a long history of rule by absolute central authority. As a result, there is a tendency to obey and submit to that authority without questioning its legitimacy or its fairness. Such a strong social influence colors not only OPFOR thinking about themselves but thinking about FFOR.

The psychological means of deception relating to cultural behavior involves exploiting the "blinders" that any culture places on the thought processes of individuals. Based on the role of culture in guiding explanation, these blinders provide a model of reality that each culture and subculture use to interpret and explain events. It predisposes them toward certain kinds of explanations and away from others. An understanding of the culturally preferred modes of explanation gives a FFOR deception planner a way of constructing a plan that will appear believable to a given OPFOR. It also gives a way of anticipating facts and events that might not appear believable to the OPFOR, even though they are factual.

Personal influences play a role as a subcomponent of behavioral means. No individual in society - even a society that prizes conformity - is strictly a product of cultural influences. Personal experiences also exert a strong influence on thinking

Table 1. Individual Information Processing Means.

Overload Short Term Memory Limits:	It has been well-accepted that humans have severe limits on the number of "chunks" of new information that can be stored in their short term, or working memory. This limitation suggests an avenue for deception -- overloading the short term memory of the targeted OPFOR decision-maker. In practice, they may be difficult to do, but not impossible. Simply increasing the amount of information that the decision-maker must manipulate can strain this limitation and lead to altered decision behavior
Overload Information Processing Rate Limits:	This factor, related to the previous one, refers to limits in the *speed* at which humans can perceive, encode, interpret, and manipulate information. The deception avenue posed by this limitation is to increase the rate at which events, data, and information reach the attention of the targeted decision maker. This can work particularly well in combination with the preceding avenue which suggests increasing the data volume.
Appeal to the Law of Small Numbers:	People will often make conclusions or interpretations that are not justifiable statistically about the generality of some phenomenon based on only a small number of instances. They generalize too quickly. The clear deception means presented by this aspect of information processing is to deliberately create a small number of data points that support the generation of a specific "picture" that the planner wants the OPFOR commander to "see".
Apply Conditioning:	This bias refers to the difficulty people have in detecting small increments in indicators, even if the cumulative change is significant. It likely involves a strong interaction between expectation and perception, as people are notorious for perceiving what they expect to happen based on interpretations of very recent perceptions. The psychological means involved here is that of using this tendency to move the targeted decision maker slowly but consistently toward a desired perception.

Table 1. Individual Information Processing Means. (Continued)

Assume Inability to Integrate Information:	The deception means presented by this limitation lies in knowing what information features the decision maker gives great emphasis. This can be used to induce a perception that is totally inconsistent with the overall statistical picture.
Exploit the Inability to Deal With Negative Information:	Another side of the link between expectation and perception is that people systematically *fail* to seek or appreciate the significance of evidence which controverts their preconceptions, models, and views of the world. In addition, they fail to seek simple-to-obtain data that could disprove their hypotheses. This bias sems to be the underlying rationale for Magruder's Principles, cited in draft FM 90-2: "It is generally easier to induce an enemy to maintain a preexisting belief than to present notional evidence to change that belief". This deception means builds on the fact that once specific hypotheses or perceptions of the OPFOR can be identified (or better still, induced), they can be easily maintained through provision of confirmatory information.

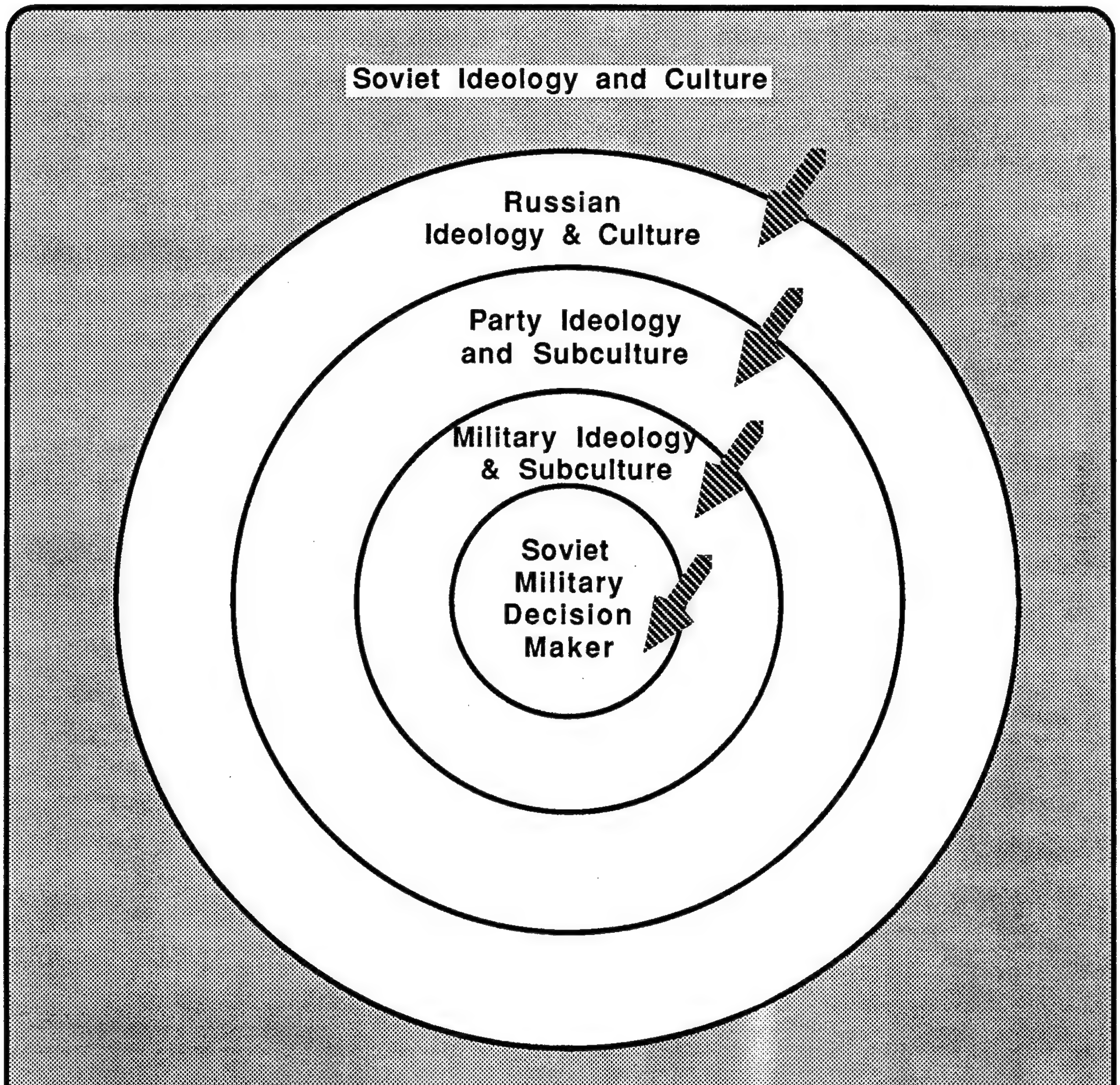


Figure 4. Cultural Influences on Soviet Military Decision-Making

and actions. In complicated interaction with the more general cultural variables, personal history helps to create an individual system of preferences (e.g., for particular tactical maneuvers, personnel, and materiel use), and style. This information about the specific OPFOR commander or decision maker facing the FFOR can be extremely valuable to the deception planner. If obtainable, the answers to questions such as, how has he thought and acted in recent tactical situations, and, what were the outcomes of these situations, can allow the development of a deception plan which goes along very precisely with what the particular individual is predisposed to believe.

Individual affective means. The affective category of individual psychological means refers to inducing emotional reactions via tactical deception. Such reactions are typically considered to be in the province of psychological operations rather than deception, but are considered here because they affect the perceptions and actions of the tactical decision maker. Actions that produce strong emotional reactions such as fear can inhibit reasoned, rational problem solving and decision making. This may be especially significant for the Soviet OPFOR because of their reliance on "algorithmic rationality". Confusion, uncertainty, and surprise are affective responses which can aid the FFOR operations by inducing the OPFOR to abandon the preplanned, algorithmically confined course of action. Often the effect of a strong emotional jolt is to just stop all activity temporarily.

An important affective response class is stress. Stress is a very complex group of variables, arising from a number of diverse types of conditions. Some of these are: time pressure, equipment failure, injury, deprivation of sleep or food, as well as confusion and fear. Stress may be assessed both in terms of conditions which are conductive to its presence and the performance decrements which accompany high levels of stress. As stress increases, one kind of performance decrement commonly seen is reduction in mental effectiveness. Particularly prominent among these are the abandonment of recently learned behaviors and the overemphasis of ingrained, older patterns, even if these are highly inappropriate from a decision perspective. Stress should be used not as a single means for deception, but in conjunction with another action, e.g., a notional or actual attack. Stress may be a particularly useful means to achieve the goals of confusing the enemy or gaining time.

Organizational psychological means. The second major class of psychological means are the organizational means depicted in the right half of the diagram of Figure 3. These influences, in contrast to the individual means, act on the force decision maker through the entire organization. That is, organizational psychological means are derived from characteristics and vulnerabilities of the organizational contexts of individual decision makers. In discussing these, an attempt has been made to utilize the same terminology as that in the individual branch

(information processing, behavioral, affective), even though this necessitates a somewhat broader meaning than that usually associated with the terms.

Organizational information processing means. This category is analogous to the individual means. The information processing structure of an organization refers to that structure - functional components and their interrelationships - which is involved with the organizational processing of information. This processing includes obtaining, interpreting, storing, and manipulating information for the purpose of making and implementing command decisions. Different organizations have different processing structures to accomplish their goals and objectives. These structures are sometimes explicitly designed to accomplish the goals; sometimes the structures evolve "naturally", without specific plans. Nevertheless, each information processing structure has certain strong points and vulnerabilities which derive directly from the structure itself. Five means of exploiting organizational information processing characteristics are listed and defined in Table 2.

Organizational behavioral means. The second class of organizational psychological means refers to exploitable characteristics that are learned rather than inherent in the structure. The organizational behavioral factors are cultural, small group processes, and formal vice informal processes.

There are a number of concurrent cultural influences on any individual which stem from the different cultures of which that individual is a member. In the case of the Soviet OPFOR, the military, the Communist Party, the Soviet and Russian cultures all influence the thinking and behavior of a given individual. In the discussion of individual means, all these influences were seen as affecting just the commander or tactical decision maker; in the organizational context, these layers of culture influence the interactions within the commander's organization. Their important effect in this context is to circumscribe the areas and forms of discourse that can occur, or at least that can be openly considered.

Small group processes can be exploited by deception. A commonly known group dynamic is "groupthink". This is the process by which a group of individuals becomes isolated from and impervious to feedback from the outside world. Their thinking becomes conformist and stylized, and therefore susceptible to deceptions that reinforce this thinking. When groupthink is operating, all members of the organization view the world in the same way and confirm each other's interpretations of events rather than seeking to disprove or critically evaluate them. The means of exploiting this phenomenon, if it can be detected, is to construct a "story" that the command group wants to believe, and then to consistently reinforce that perception in an attempt to create a groupthink acceptance of it.

Table 2. Organizational Information Processing Means.

Degree of Centralization:	This refers to the extent to which key information processing decisions are made by a central, high-level authority or are distributed and made in parallel. Key decisions here include both sensing/interpreting of information and planning/implementing actions. Highly centralized information processing structures are better able to achieve consistency of response and efficiency of decision making. They are also, however, relatively insensitive to distant localized conditions and unable to rapidly respond to unexpected stimuli. Distributed structures are less able to achieve unified and tightly coordinated action, but are more able to perceive and respond rapidly to local conditions and unexpected stimuli. Both structures pose opportunities for deception.
Degree of Preplanning:	This characteristic refers to the extent to which specific intelligence and operations activities are planned in detail before the engagement begins. Advantages or preplanning are speed and consistency of response; the primary disadvantage is confusion if unanticipated contingencies occur. Organizations with lower degrees of preplanning are initially slower to act. However, they have no anticipations already implicitly built into pre-existing plans and they are more able to act quickly with a broader range of options.
Formalization of Decision Making:	This feature is concerned with the extent to which planning and decision-making are based on formal, standardized methods (e.g., models, formulas, simulations, nomograms) or on individual decision maker criteria and experiences. When standardization and formalization are high, the organization achieves greater consistency, formality, and "rationality". This also makes its behavior more predictable, and can reduce creativity and learning on the part of individual decision makers. Use of individual criteria, on the other hand, can lead to decisions that are inconsistent across units of the organization and problems in coordination of decisions. Individual criteria allow more flexible responses and permit development of more expertise on the part of key decision makers.

Table 2. Organizational Information Processing Means. (Continued)

Technological Style:	Each command organization will have a general technological style or approach. There may be reliance on: high-tech equipment; preferences for certain types of sensors or weapons; man-machine function allocation; or a general avoidance of such reliance. The technological reliance leads to improved capabilities under nominal conditions, but may lead to degraded performance in highly stressed problematic situations. An avoidance of technological reliance may make the organization more resistant to stress-induced failure, but also generally slows the speed of and capacity for information flow in the organization.
Organizational (Command and Control) Structure:	The nature of the C2 structure -- what functions are linked and how, the degree of redundancy, etc. -- is an important determinant in the vulnerabilities of a force. Highly redundant systems have increased chances of successful communication, but typically allow a smaller variety of messages. Such a system may be susceptible to: overconfidence, if several channels are simultaneously deceived; or confusion, if divergent information is sent to distinct channels. Less redundant systems are susceptible to the loss of information or the loss of functionality if key communication links or organizational components are lost.

The distinction between formal and informal behavioral organizational means involves detecting the explicit and implicit ways in which an organization operates. Military organizations are highly and explicitly organized and focus on formal entities, such as structures and processes established via organizational charts, procedural manuals, etc. Nevertheless, informal structures and processes exist in all organizations that have been studied, including the military. These informal structures represent the organization's responses to idiosyncratic conditions or individual patterns of behavior, as well as informal channels of communication. For example, each military organization has a formal command structure, from high to low (e.g. Army to squad) levels. In theory all decisions flow downward from the highest levels of this chain to the lowest. Yet it is known that there is a network of individuals who informally pass information and/or initiate decisions in the opposite direction, from the bottom to the top, and bypass established channels. This network is just as real as the formal chain of command, yet it is what is called informal.

Under conditions of stress, formal structures tend to become simplified, while informal structures and processes tend to become more elaborated, as individuals develop ad hoc means of circumventing stressing conditions. One way in which deception can be introduced is to place information into the informal network that is contradictory to information in the formal system. This can provide a confusing effect, as commanders receive different conclusions from the two sources. In fact, deception can enhance the conflicts in a military organization that stem from formal and informal discrepancies. In the Soviet system, the "matrix" organization of tactical and political command often has informal rivalry as well as formal cooperation.

Organizational affective means. The final category of organizational means refers to ways to induce disruptive or destructive emotional reactions into a tactical organization. Psyops is appropriately placed in this category, as in the individual branch.

Organizational affective factors deal with the relationships among individuals within an organizational context. Using deception to introduce negative and distracting emotional dynamics into these relationships - including fear, anger, distrust, jealousy, despair - can be effective disrupting operations.

Organizations can also be stressed just as individuals can. In the organization, stress relates to coordinative, cooperative, integrative activities; those that require interrelationship of roles and tasks. Building upon or enhancing naturally volatile dynamics would seem to be an effective way to utilize stress as a deceptive means. This approach also draws upon the formal vice informal distinction above, and the rivalries that often accompany it. Similarly, accentuating the competition for limited

resources (e.g., multiple simultaneous attacks) can cause organizational stress and performance decrements.

The formal vice informal distinction applies to affective means as well as the behavioral since it can be used to create conflict and distrust, and disrupt the emotional well being of the enemy force. Exploiting the discrepancies between formal and informal networks is an important deception means which may be behaviorally or affectively directed, but may be most effective when directed at both.

Operational Means for Deception

Operational means for deception is the second branch of FFOR exploitation means diagramed in Figure 1. This part of the means structure concerns the operational and tactical action, information control measures, and physical resources which might be used in conjunction with the psychological means to deceive the OPFOR decision maker(s). Figure 5 shows a detailed branching of operational means into three categories: operational/tactical actions, information denial/control measures, and physical resources.

Operational/Tactical means and information denial/control measures. Operational/tactical actions refer to the observable activities which compose the planning, preparation and execution of courses of action at the operational sector and/or tactical levels. These activities are usually not deceptive in nature, but may be very effectively used toward such ends. Information denial/control measures refer to the FFOR capabilities available to the deception planner which allow some control of information available to OPFOR intelligence-gathering assets. Examples of information/denial control measures include:

- standard OPSEC procedures for denying information

- intentionally breaking OPSEC on a notional operation

- use of deception specific resources to portray a false situation

- tight procedural control over visual, olfactory, sonic, thermal and electronic/communications emissions by an armored cavalry squadron as it lies in ambush along a route the OPFOR is using in response to false indicators of the strength and deployment of units in the area

- use of expended resources, such as irreparably damaged parts which appear to be new or under repair, to simulate a direct support maintenance unit.

Denial/control measures can best be categorized by the sensory channel through which the OPFOR obtains the information. Thus as shown in Figure 5, these measures may be classified into

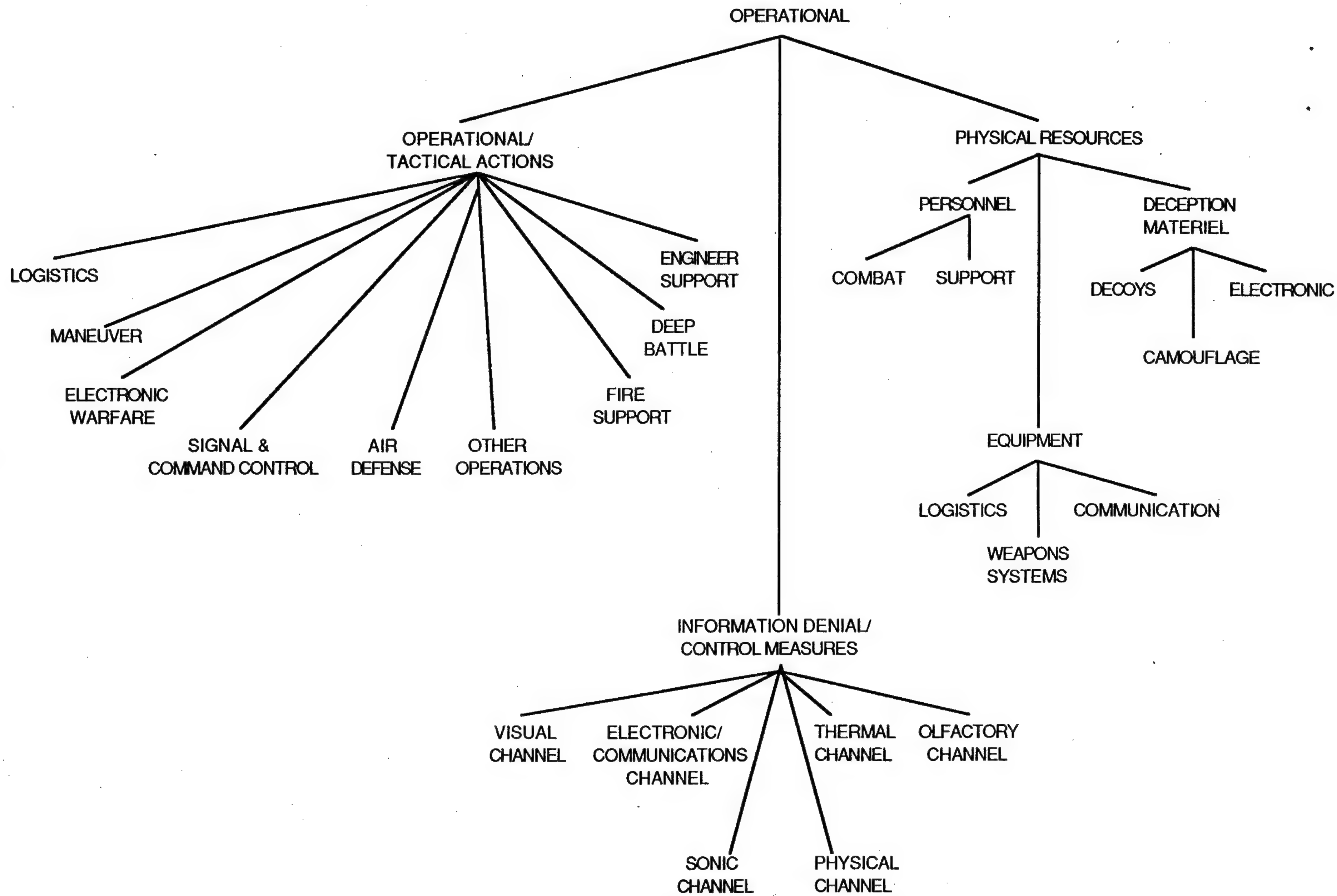


Figure 5. Operational Means Component

measures affecting the visual channel, electronic/communications channel, sonic channel, etc. Information denial measures are procedures or actions applied to applicable sensory channels to inhibit the flow of salient information, while information control measures are procedures or actions applied to applicable sensory channels to regulate the quantity, nature, form, timing, and selection of channels to the OPFOR.

Physical resource means. The third branch of operational deception means is based upon specific physical resources. Physical resources can be used in a variety of specific ways in deception operations to support or link together the larger pieces of the plan. Figure 5 depicts the three sub-branches of physical means as personnel, equipment, and deception materiel.

Personnel. Personnel are the military commander's most flexible and valuable asset. The proper employment of soldiers on the battlefield is of critical importance, whether being used as part of the main defensive or offensive maneuver, or as part of a deception designed to lead the OPFOR commander into believing a false or misleading scenario. Personnel may be taken away from their more typical primary tasks to perform deception tasks. Their appropriate employment for deception must be more beneficial to the total mission than their loss to their primary functions is detrimental.

Equipment. Materiel that is organically assigned to the unit as part of its TO&E for purposes other than deception is referred to as "equipment". In most cases, some military equipment will be required as part of any deception activity. The equipment might be sophisticated weapons systems, being employed to create the perception of an attacking or defending force where there really is none. It might also be as simple as truck or personal equipment and weapons, being used to portray a platoon or company headquarters. Even expended, damaged, and unrecoverable equipment can be a deception resource. For example, a forward area supply point can be simulated using empty fuel bladders, and expended ammunition boxes and casings. These items give the illusions of activities or troop concentrations without tying up valuable assets.

Deception materiel. The combat commander may have available an array of specifically designed equipment for deception purposes. Ranging in sophistication, this equipment includes inflatable decoys which can portray artillery or air defense weapons, as well as sophisticated "black boxes" which simulate electronic and mechanical environmental signatures. For example, by providing appropriately camouflaged dummy equipment and a black box to simulate infrared (IR) and other electronic signatures, a near complete decoy package may be constructed without committing any actual weapons systems to deception activities. As this type of deception equipment becomes more readily available, the range of deception options will increase.

Since the OPFOR commander will rarely rely solely on any one source of intelligence for decision making, realism, both in procedures and physical appearance is a critical feature of the deception story. For this reason, effective use of personnel, deception materiel, and other equipment in combination is necessary to simulate any specific situation.

OPFOR Decision Cycle Vulnerabilities Component

The preceding branches of the tactical deception framework have defined deception goals or objectives, and the different types of means available that can be brought to bear in order to achieve the specific goal. The final framework branch (see Figure 1) identifies the components which must be considered in developing and evaluating different deception strategies. This requires an examination of the OPFOR decision cycle by considering vulnerabilities related to decision makers, timing constraints, and susceptibility of the OPFOR to a particular plan. These branches of the vulnerabilities component are diagrammed in Figure 6.

Decision makers and organizational structure

In order to fashion a specific plan in a particular area of operations/area of interest, the specific Order of Battle data and unit line and block charts of the OPFOR must be obtained. Beyond this, however, specific data relating to the actual decision makers, formal and informal means of communication, and information flow pathways must be identified. A network representation could be used to capture the information flow (paths of the network) and to identify the actual personnel who act on that information (nodes within the framework) as well as specify their decision making roles (function of a node). Combining this network knowledge with data regarding the personnel characteristics and decision making characteristics within the command chain improves the likelihood that the deception means selected for the operational situation will be successful.

Timing Constraints

Baseline data regarding mission planning and execution phases of the OPFOR need to be integrated into estimates of the tactical situation. Although it may be difficult to plot the exact timing of specific decision and execution phases for a given battlefield period, current intelligence should provide indicators of when opportune times for deception inputs exist, and what time windows are available in which to execute various tactics. As shown in Figure 6, the derivation and representation of time windows involves detection of physical, situational, and decision cycle pathways and flows. Physical timing constraints are those associated with movement, fires, communication, and resupply, and are based on the time it takes to carry out these actions. Situational timing constraints are either tactical or

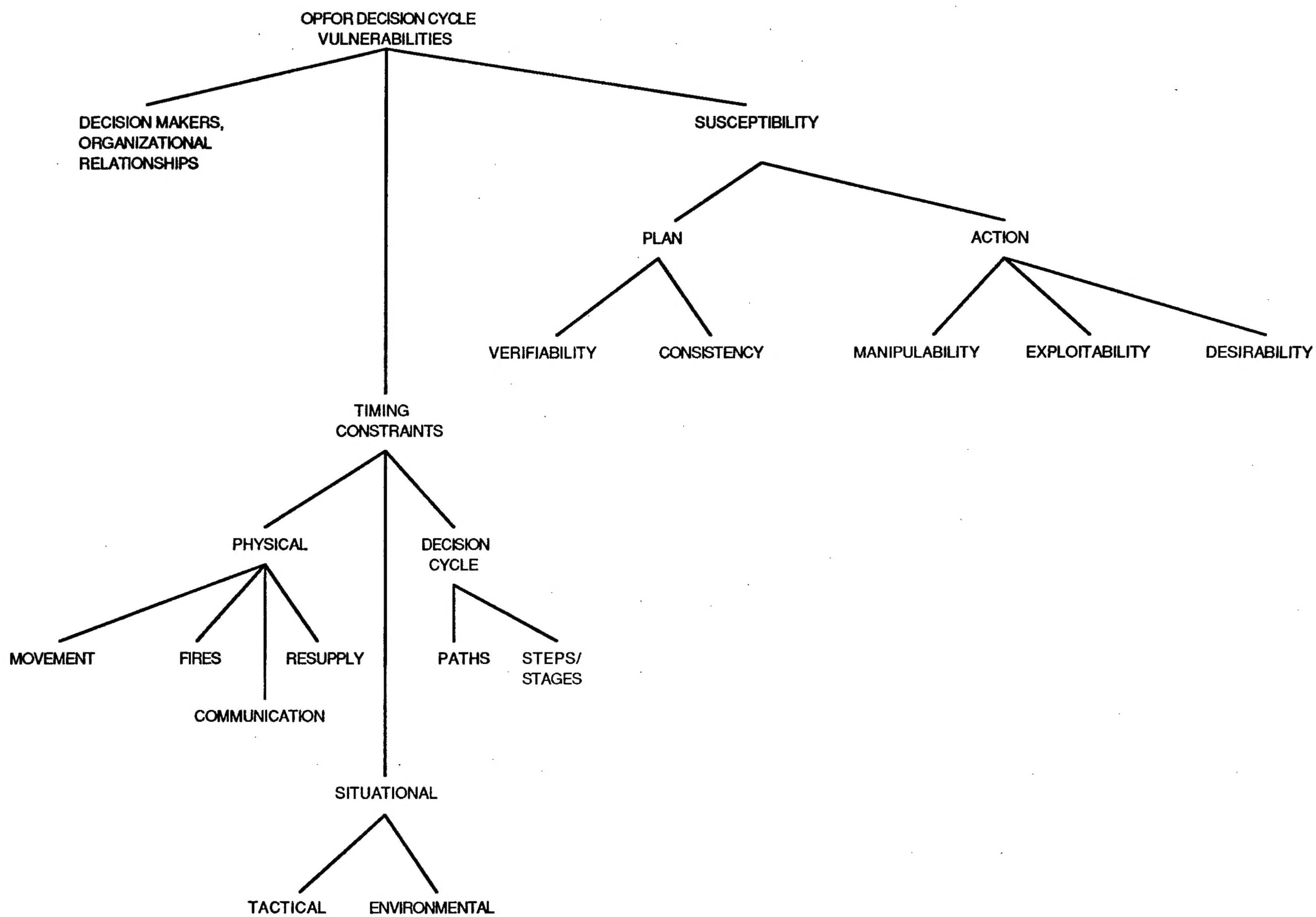


Figure 6. OPFOR Decision Cycle Vulnerabilities Component

environmental, and include the time required to implement and execute tactical actions, and the time required to act under environmental conditions that prevail (terrain, weather, etc.). Decision cycle constraints are associated with the paths and steps in the OPFOR decision cycle, which includes time to pass information, times to act on that information, as well as communication mode used.

Susceptibility

An examination of susceptibility of the OPFOR to deception involves reviewing two perspectives of the plan: the FFOR capability to react and execute the plan, and an OPFOR view of the plan in terms of whether plan execution will be compatible with the unfolding OPFOR decision cycle.

In the FFOR view, a wargaming process must ensue which examines whether the planner has in fact identified the right information pathways, timing windows, and execution measures which will cause plan success. Important questions to be resolved are whether the timing windows can be met, and whether information flows can be controlled throughout the deception operation in order to insure the plan will not be exposed. The war-gaming process is required in order to visualize the action/reaction process on the part of each force that will occur, and in what approximate sequence and time frames, in order to determine the likelihood of plan success. This is an analysis process that assesses the deception strategy chosen purely in terms of the FFOR capability to react to OPFOR tactics taken in the course of the operation.

From the OPFOR view, the susceptibility component of deception is also a final assessment of the weak links in the OPFOR decision cycle from the OPFOR perspective. Since goals and means have now been matched with detected decision maker vulnerabilities and time windows, and war-gamed according to FFOR capability to react and execute the plan, a final check is needed for assessing the overall consistency of the plan, and a means established to monitor success of its execution (verifiability). The issue to be considered in this second view of susceptibility is to determine whether the deception plan is compatible with current intelligence as represented in the planning process thus far. Essentially the entire plan can now be viewed according to desirability of the plan as a function of manipulability and exploitability of the OPFOR. Although somewhat abstract in nature, these terms imply a process whereby the goals, means, nodes, paths, and time windows examined in constructing the deception plan will actually work under varying conditions. Desirability analysis examines whether affecting the decision cycle will lead to accomplishing the FFOR commander's mission. This gives a measure of the degree to which the OPFOR action, taken as a result of the deception operation, will actually benefit the FFOR. The deception strategist must take a systematic approach to assessing the plan in light of the most recent

operational plans. Manipulability examines the extent that the FFOR can change OPFOR action by looking at what is actually sensitive, that is, what channels are important to the OPFOR, and how are they used. Finally exploitability examines the extent to which the FFOR plan can actually affect the selected information paths or input channels as they function in the OPFOR system.

CONCLUSIONS

The preceding description of a framework for tactical deception points out structures, concepts and entities, along with their definitions, in order to organize what needs to be considered in the battlefield deception domain. The three part framework consists of: 1) a definition of goals, both in terms of the FFOR operational mission and desired OPFOR actions in relationship to those goals, followed by 2) an elaboration of psychological and operational means available to meet the goals, and 3) vulnerability analyses of the OPFOR decision cycle with its nodes, links, and decision makers, required in order to effectively employ the selected deception strategy.

The deception framework, as presented, serves as a high level "checklist" of the entities to be considered in the tactical deception domain. It allows determination as to whether all relevant dimensions of the deception domain have been considered. In its present form, it is a precursor to further research which will elaborate the dynamic interrelationships between goals and means, as well as the mechanics of the analyses required to assess vulnerability for a particular deception operation. This further effort involves operationalizing the concepts and structures into a form that can be used to aid a deception planner, developing appropriate psychological and operational knowledge bases for use by the planner, and employing modelling techniques to capture the dynamics of battlefield situations and decision cycles in order to perform the analyses required.